

Intertwinings and Stein's magic factors for birth-death processes

Bertrand Cloez*

Claire Delplancke†

September 28, 2016

Abstract

This article investigates second order intertwining between semigroups of birth-death processes and discrete gradients on \mathbb{N} . It goes one step beyond a recent work of Chafaï and Joulin which establishes and applies to the analysis of birth-death semigroups a first order intertwining. Similarly to the first order relation, the second order intertwining involves birth-death and Feynman-Kac semigroups and weighted gradients on \mathbb{N} , and can be seen as a second derivative relation. As our main application, we provide new quantitative bounds on the Stein factors of discrete distributions. To illustrate the relevance of this approach, we also derive approximation results for the mixture of Poisson and geometric laws.

Keywords: Birth-death processes; Feynman-Kac semigroups; intertwining; Stein's factors; Stein's method; distances between probability distributions.

Mathematics Subject Classification (MSC2010): 60E15; 60J80, 47D08, 60E05, 60F05.

Contents

1	Introduction	2
2	Main result	4
3	Application to Stein's magic factors	9
3.1	Distances between probability distributions	9
3.2	Basic facts on Stein's method	11
3.3	Bounds on Stein's magic factors	12
3.4	Stein's method and mixture of distributions	15
4	Examples	17
4.1	The M/M/ ∞ process and the Poisson approximation	17
4.2	The GWI process and the negative binomial approximation	19
4.3	The M/M/1 process and the geometric approximation	23
4.4	Another example	26
5	Proofs of Section 2	26
5.1	First order intertwining for the backward gradient ∂_u^*	26
5.2	Alternative proof of first order intertwining theorems	27
5.3	Proof of Theorems 2.2 and 2.5	29

*UMR MISTEA, INRA Montpellier, France.

†UMR CNRS 5219, Institut de Mathématiques de Toulouse, Université de Toulouse, France.

E-mail address: claire.delplancke at math.univ-toulouse.fr (corresponding author).

6	Proofs of Section 3	30
6.1	Approximation in total variation distance	31
6.2	Approximation in Wasserstein distance.	32
6.3	Approximation in Kolmogorov distance	34
7	Proof of Section 4	36

1 Introduction

A birth-death process is a continuous-time Markov process with values in $\mathbb{N} = \{0, 1, \dots\}$ which evolves by jumps of two types: onto the integer just above (birth) or just below (death). We denote by $\text{BDP}(\alpha, \beta)$ the birth-death process with positive birth rate $\alpha = (\alpha(x))_{x \in \mathbb{N}}$ and non-negative death rate $\beta = (\beta(x))_{x \in \mathbb{N}}$ satisfying to $\beta(0) = 0$. Its generator is defined for every function $f : \mathbb{N} \rightarrow \mathbb{R}$ as

$$Lf(x) = \alpha(x)(f(x+1) - f(x)) + \beta(x)(f(x-1) - f(x)), \quad x \in \mathbb{N}.$$

For a generator L , associated to a semigroup $(P_t)_{t \geq 0}$ and a Markov process $(X_t)_{t \geq 0}$ on \mathbb{N} , and a function V on \mathbb{N} (usually called a potential), the Schrödinger operator $L - V$ is defined for every function f as

$$(L - V)f(x) = (Lf)(x) - V(x)f(x), \quad x \in \mathbb{N},$$

and is associated to the Feynman-Kac semigroup $(P_t^V)_{t \geq 0}$ defined for all bounded or non-negative functions f on \mathbb{N} as

$$(P_t^V f)(x) = \mathbb{E} \left[f(X_t^x) e^{-\int_0^t V(X_s^x) ds} \right], \quad x \in \mathbb{N}, \quad t \geq 0.$$

The starting point of our work is the recent article Chafaï and Joulin [2013] which establishes a first order intertwining relation involving birth-death and Feynman-Kac semigroups, and discrete gradients on \mathbb{N} . For example, it reads as

$$\partial P_t = \tilde{P}_t^{\tilde{V}} \partial, \tag{1}$$

where ∂ is the discrete gradient defined by $\partial f(x) = f(x+1) - f(x)$, the notation $(\tilde{P}_t^{\tilde{V}})_{t \geq 0}$ standing for an alternative Feynman-Kac semigroup. Actually, the precise result holds more generally for weighted gradients and allows to derive known as well as new results on the analysis of birth-death semigroups.

According to this observation, the aim of the present article is to extend this work by stating a second order intertwining relation. More precisely, let us define the backward gradient ∂^* by

$$\partial^* f(x) = f(x-1) - f(x), \quad x \in \mathbb{N}^* = \{1, 2, \dots\}; \quad \partial^* f(0) = -f(0).$$

Under some appropriate conditions on the potential \tilde{V} , we derive a formula of the type

$$\partial^* \partial P_t = \hat{P}_t^{\hat{V}} \partial^* \partial, \tag{2}$$

where $(\hat{P}_t^{\hat{V}})_{t \geq 0}$ is a new Feynman-Kac semigroup. Similarly to the first order, this second order intertwining relation, which is our main result, is given in the more general case of weighted gradients.

Once our second order relation is established, it reveals to have many interesting consequences. In particular, we derive results on the estimation of the so-called Stein factors. Stein's factors,

also known as Stein's magic factors, are upper bounds on derivatives of the solution to Stein's equation and a key point in Stein's method, introduced by Stein in Stein [1972], which consists in evaluating from above distances between probability distributions. Among the important results appearing more or less recently in this very active field of research, let us cite some references within the framework of discrete probabilities distributions. Stein's factors related to the Poisson approximation in total variation and Wasserstein distances are studied in the seminal paper Chen [1975], in the reference book Barbour et al. [1992] and in the recent article Barbour et al. [2015] for example. For the binomial negative approximation, one can cite Brown and Phillips [1999] for the total variation distance and Barbour et al. [2015] for the Wasserstein distance; for the geometric approximation in total variation distance, see Peköz [1996] and Peköz et al. [2013]. An important advance is made in Brown and Xia [2001], where a universal approach to evaluate Stein's factors for the total variation distance is developed. The work Eichelsbacher and Reinert [2008] provides Stein's factors for the total variation distance when the target distribution is a Gibbs distribution.

In the present article, we propose a universal technique to evaluate Stein's factors related to the approximation in total variation, Wasserstein and Kolmogorov distances. On the basis of some results derived in Brown and Xia [2001], the main ingredients are the so-called method of the generator and the intertwining relations presented above. To the authors' knowledge, the systematic use of this last ingredient, which comes from the functional analysis, seems to be new within the context of Stein's method. It allows to construct a unified framework for the derivation of Stein's factors, which applies to a wide range of discrete probability distributions, and might be developed similarly in the continuous setting of diffusion processes in a forthcoming paper.

As a result, a case-by-case examination has to be done in each situation and example of interest, revealing that our upper bounds sometimes improve on the ones already known, and sometimes are not as sharp. For example, we improve the first Stein factor related to the negative binomial approximation in total variation distance and we derive new Stein's factors for the geometric approximation in Wasserstein distance.

As an additional part of independent interest, we study the approximation of mixture of discrete distributions in the spirit of the Stein method. Combined with the Stein bounds, the obtained results have potential applications of which we give a flavour through the following example. Denote $\text{NB}(r, p)$ the negative binomial distribution of parameters (r, p) . It is a mixed Poisson distribution, converging in law towards the Poisson distribution \mathcal{P}_λ in the regime $p \rightarrow 1$, $r \rightarrow \infty$ and $r(1-p)/p \rightarrow \lambda$. The following bound in Wasserstein distance W seems to be the first attempt to quantify this well-known convergence:

$$W\left(\text{NB}(r, p), \mathcal{P}_{\frac{r(1-p)}{p}}\right) \leq \frac{8}{3\sqrt{2e}} \sqrt{\frac{r(1-p)}{p} \frac{(1-p)}{p}}.$$

To conclude this introduction, let us announce the structure of the article. In Section 2, we state with Theorem 2.2 our main result about the second order intertwining, after having recalled the first order intertwining; we follow with an application to the ergodicity of birth-death semigroups. In Section 3, we firstly present theoretical bounds on Stein's factors derived from the intertwining, and secondly we investigate the approximation of mixture of distributions. In Section 4, our results are applied to a wide range of examples, including M/M/ ∞ process and Poisson approximation, Galton-Watson process with immigration and negative binomial approximation, and M/M/1 process and geometric approximation. The three last sections are devoted to the various proofs of the results previously stated: Section 5 deals with the preparation and proof of our main result Theorem 2.2, Section 6 gathers the proofs of the bounds on Stein's factors

and finally, a useful upper bound related to the pointwise probabilities of the $M/M/\infty$ process is proved in Section 7.

Acknowledgement:

The authors thank A. Joulin for the interest he took in this research through many fruitful discussions and for sharing his insights on the subject.

This work was partially supported by the CIMI (Centre International de Mathématiques et d'Informatique) Excellence program, by the ANR PIECE (ANR-12-JS01-0006-01) and STAB (ANR-12-BS01-0019) and by the Chaire Modélisation Mathématique et Biodiversité.

2 Main result

Before stating our main result Theorem 2.2, let us introduce some notation. The set of positive integers $\{1, 2, \dots\}$ is denoted \mathbb{N}^* . For all real-valued functions f on \mathbb{N} and sets $A \subset \mathbb{N}$, we define $\|f\|_{\infty, A} = \sup\{|f(x)|, x \in A\}$ and $\|f\|_{\infty} = \|f\|_{\infty, \mathbb{N}}$. For all sequences u on \mathbb{N} , the shift-forward and shift-backward of u are defined as:

$$\vec{u}(x) = u(x+1), \quad x \in \mathbb{N}; \quad \overleftarrow{u}(x) = u(x-1), \quad x \in \mathbb{N}^*; \quad \overleftarrow{u}(0) = 0.$$

The symbol \mathcal{P} stands for the set of probability measures on \mathbb{N} and we denote by $\mathcal{L}(W)$ the distribution of the random variable W . For all real-valued functions f on \mathbb{N} and $\mu \in \mathcal{P}$, we use indifferently the notation

$$\int f d\mu = \mu(f) = \sum_{x \in \mathbb{N}} f(x) \mu(x).$$

Recall that the discrete forward and backward gradients are defined for all real-valued functions f on \mathbb{N} by

$$\partial f(x) = f(x+1) - f(x), \quad x \in \mathbb{N}; \quad \partial^* f(x) = f(x-1) - f(x), \quad x \in \mathbb{N}^*, \quad \partial^* f(0) = -f(0),$$

the convention chosen for ∂^* in 0 being interpreted as a Dirichlet-type condition (implicitly we set $f(-1) = 0$). Letting u be a positive sequence, we define the weighted gradients ∂_u and ∂_u^* respectively by

$$\partial_u = \frac{1}{u} \partial, \quad \partial_u^* = \frac{1}{u} \partial^*.$$

With this notation, the generator of the $BDP(\alpha, \beta)$ reads for every function $f : \mathbb{N} \rightarrow \mathbb{R}$ as

$$Lf = \alpha \partial f + \beta \partial^* f.$$

Let us assume that the birth rate α is positive on \mathbb{N} and that the death rate β is positive on \mathbb{N}^* with moreover $\beta(0) = 0$. Hence the process is irreducible; to ensure that the process is ergodic and non-explosive we further assume respectively that (Chen [2004])

$$\sum_{x=1}^{+\infty} \frac{\alpha(0)\alpha(1)\dots\alpha(x-1)}{\beta(1)\beta(2)\dots\beta(x)} < \infty, \quad \sum_{x=1}^{\infty} \left(\frac{1}{\alpha(x)} + \frac{\beta(x)}{\alpha(x)\alpha(x-1)} + \dots + \frac{\beta(x)\dots\beta(1)}{\alpha(x)\dots\alpha(0)} \right) = \infty.$$

The measure π defined on \mathbb{N} as

$$\pi(0) = \left(1 + \sum_{x=1}^{\infty} \prod_{y=1}^x \frac{\alpha(y-1)}{\beta(y)} \right)^{-1}, \quad \pi(x) = \pi(0) \prod_{y=1}^x \frac{\alpha(y-1)}{\beta(y)}, \quad x \in \mathbb{N}, \quad (3)$$

is then the invariant, and symmetric, probability measure for the associated semigroup.

Recall that if $(P_t)_{t \geq 0}$ is a Markov semigroup on \mathbb{N} associated to the process $(X_t)_{t \geq 0}$ and if the potential $V : \mathbb{N} \rightarrow \mathbb{R}$ is bounded from below, the Feynman-Kac semigroup $(P_t^V)_{t \geq 0}$ is defined for all bounded or non-negative functions f on \mathbb{N} as

$$(P_t^V f)(x) = \mathbb{E} \left[f(X_t^x) e^{-\int_0^t V(X_s^x) ds} \right], \quad x \in \mathbb{N}, \quad t \geq 0. \quad (4)$$

When V is positive, the formula (4) admits an interpretation involving a killed, or extended, Markov process. Add a new state a to \mathbb{N} and extend functions f on \mathbb{N} to $\mathbb{N} \cup \{a\}$ by $f(a) = 0$. Then, we have:

$$P_t^V f(x) = \mathbb{E} \left[f(Y_t^x) \mathbf{1}_{\{Y_t^x \neq a\}} \right],$$

where the process $(Y_t^x)_{t \geq 0}$ is absorbed in a with rate $V(Y_t^x)$. The generator of the process $(Y_t^x)_{t \geq 0}$ acts on real-valued functions on $\mathbb{N} \cup \{a\}$ by the formula

$$(Kf)(x) = (Lf|_{\mathbb{N}})(x) + V(x)(f(a) - f(x)). \quad (5)$$

This interpretation can be extended to the case where V is bounded from below by adding and subtracting a constant to V inside the exponential.

The Kolmogorov equations associated to the Schrödinger operator $L - V$ and the Feynman-Kac semigroup defined in the introduction read for all functions f in the domain of L as

$$\partial_t P_t^V f = (L - V)P_t^V f = P_t^V (L - V)f, \quad t \geq 0. \quad (6)$$

Here ∂_t denotes the derivative in time. In the following, when using a Feynman-Kac semigroup, we will always assume that the equation (6) stands for all bounded real-valued functions on \mathbb{N} . It is the case for example when L is the generator of a birth-death process with rates (α, β) , and α, β, V are P_t -integrable for all $t \geq 0$.

In order to state the first intertwining relation, we associate to any positive sequence u a modified birth-death process on \mathbb{N} with semigroup $(P_{u,t})_{t \geq 0}$, generator L_u , and potential V_u . For all functions $f : \mathbb{N} \rightarrow \mathbb{R}$ set

$$\begin{aligned} L_u f &= \alpha_u \partial f + \beta_u \partial^* f, & V_u &= \alpha - \alpha_u + \overset{\rightarrow}{\beta} - \beta_u, \\ \alpha_u(x) &= \frac{u(x+1)}{u(x)} \alpha(x+1), & \beta_u(x) &= \frac{u(x-1)}{u(x)} \beta(x) \mathbf{1}_{x \in \mathbb{N}^*}, & x \in \mathbb{N}. \end{aligned}$$

Under the compacted form $V_u = \partial_u \left(\overset{\leftarrow}{u} \beta - u \alpha \right)$ one can see the parallel with the analogous formulas in the diffusion setting (Bakry and Émery [1985]; Bonnefont and Joulin [2014]).

We recall now the first order intertwining relation, due to Chafaï and Joulin [2013].

Theorem 2.1 (First order intertwining relation). *If V_u is bounded from below, it holds for every real-valued function on \mathbb{N} such that $\|\partial_u f\|_\infty < +\infty$ that:*

$$\partial_u P_t f = P_{u,t}^{V_u} \partial_u f, \quad t \geq 0. \quad (7)$$

Although we will not prove this result in full generality, a new proof is proposed in Section 5.2 when the weight is $u = 1$, the birth rates α are non-increasing and the death rates β are non-decreasing. This proof is based on a coupling argument and gives a probabilistic interpretation of the semigroup (and its jump rates) in the right-hand side of (7).

We now turn to the main theorem of this article. Let u and v be positive sequences and assume that the potential V_u defined above is non-increasing on \mathbb{N} . We define a modified process on \mathbb{N}

with semigroup $(P_{u,*v,t})_{t \geq 0}$ and generator $L_{u,*v}$ as follows: for all real-valued functions f on \mathbb{N} , set

$$\begin{aligned}
(L_{u,*v}f)(x) &= \alpha_{u,*v}(x)\partial f(x) + \beta_{u,*v}(x)\partial^* f(x) \\
&\quad + (\partial_v^* V_u)(x) \left(\sum_{j=0}^{x-2} v(j) \right) \sum_{k=0}^{x-2} \frac{v(k)}{\left(\sum_{j=0}^{x-2} v(j) \right)} (f(k) - f(x)), \quad x \geq 2, \\
(L_{u,*v}f)(x) &= \alpha_{u,*v}(x)\partial f(x) + \beta_{u,*v}(x)\partial^* f(x), \quad x = 0, 1, \\
\alpha_{u,*v}(x) &= \frac{v(x+1)}{v(x)} \frac{u(x+1)}{u(x)} \alpha(x+1), \quad x \in \mathbb{N}, \\
\beta_{u,*v}(x) &= \frac{v(x-1)}{v(x)} \frac{u(x-2)}{u(x-1)} \beta(x-1) + v(x-1) \partial_v^* V_u(x), \quad x \geq 2, \\
\beta_{u,*v}(1) &= v(0) \partial_v^* V_u(1), \quad \beta_{u,*v}(0) = 0.
\end{aligned}$$

In contrast with the previous semigroups, this modified process is not a birth-death process in general. Indeed, if the process starts at a point $x \geq 2$, it can jump on the set $\{0, \dots, x-2\}$ with rate $(\partial_v^* V_u)(x) \left(\sum_{j=0}^{x-2} v(j) \right)$. Remark that both this quantity and the death rate in 1, $\beta_{u,*v}(1) = (\partial_v^* V_u)(1)$, are non-negative thanks to the hypothesis V_u non-increasing on \mathbb{N} . We also define the potential $V_{u,*v}$ as

$$\begin{aligned}
V_{u,*v}(x) &= \left(1 + \frac{u(x)}{u(x-1)} \right) \alpha(x) - \left(1 + \frac{v(x+1)}{v(x)} \right) \frac{u(x+1)}{u(x)} \alpha(x+1) \\
&\quad + \beta(x+1) - \frac{v(x-1)}{v(x)} \frac{u(x-2)}{u(x-1)} \beta(x-1) - \left(\sum_{j=0}^{x-1} v(j) \right) \partial_v^* V_u(x), \quad x \geq 1, \\
V_{u,*v}(0) &= \alpha(0) - \left(1 + \frac{v(1)}{v(0)} \right) \frac{u(1)}{u(0)} \alpha(1) + \beta(1).
\end{aligned}$$

We are ready to state our main result.

Theorem 2.2 (Second order intertwining relation). *Assume that V_u is non-increasing, bounded from below, that $\inf_{x \in \mathbb{N}} v(x) > 0$ and that $V_{u,*v}$ is bounded from below. Then for every real-valued function on \mathbb{N} such that $\|\partial_u f\|_\infty < +\infty$, we have*

$$\partial_v^* \partial_u (P_t f) = P_{u,*v,t}^{V_{u,*v}} (\partial_v^* \partial_u f), \quad t \geq 0.$$

Since some preparation is needed, the proof of Theorem 2.2 is postponed to Section 5.

Remark 2.3 (Propagation of convexity ?). *Under the assumptions of Theorem 2.2, if $\partial_v^* \partial_u f$ is non-negative, so is $\partial_v^* \partial_u P_t f$ for all $t \geq 0$. A similar property for the first order intertwining admits an interpretation in terms of propagation of monotonicity ([Chafaï and Joulin, 2013, Remark 2.4]): the intertwining relation (7) implies that if a function $f : \mathbb{N} \rightarrow \mathbb{R}$ is non-decreasing, then so is $P_t f$ for every $t \geq 0$. However, it is not clear whether there is an analogous nice interpretation for the second order intertwining because, in contrast to the continuous space case, the condition $\partial_v^* \partial_u f \geq 0$ is not equivalent to the convexity of f (even for $u = v = 1$).*

Let us comment further on Theorem 2.2. The interpretation of a Feynman-Kac semigroup as an extended Markov semigroup sheds light on various aspects of Theorem 2.2. As the first-order potential V_u is bounded from below, recall that the Feynman-Kac semigroup $(P_{u,t}^{V_u})_{t \geq 0}$ appearing in the right-hand side of equation (7) can be represented as a Markov semigroup $(S_t)_{t \geq 0}$ related to the process $(Y_t)_{t \geq 0}$ on $\mathbb{N} \cup \{-1\}$ by adding a point $a = -1$. The Markov process $(Y_t)_{t \geq 0}$ is then non-irreducible and absorbed in -1 . To differentiate again in the equation (7) amounts to differentiate the Markov semigroup $(S_t)_{t \geq 0}$.

Firstly, this explains intuitively the use of the backward weighted gradient ∂_u^* instead of the regular weighted gradient ∂_u . Indeed, to deal with the absorption of the Markov process in -1 , additional information at the boundary is needed. The use of ∂^* gives the missing information, since the knowledge of ∂^*g is equivalent to the knowledge of ∂g in addition with the knowledge of $g(0) = -\partial^*g(0)$.

Secondly, this allows to understand the hypotheses required for Theorem 2.2 to apply. The main assumption of this theorem is V_u to be non-increasing. As noticed before, this assumption is necessary in order to have well-defined objects. The following remark provides another justification.

Remark 2.4 (Around the monotonicity assumption). *On the one hand, the second intertwining relation is equivalent to a first intertwining relation for the extended Markov semigroup $(S_t)_{t \geq 0}$. On the other hand, if a first intertwining relation holds for $(S_t)_{t \geq 0}$, then $(S_t)_{t \geq 0}$ propagates the monotonicity. Set $f = \mathbf{1}_{\mathbb{N}} = 1 - \mathbf{1}_{\{-1\}}$. Then for all $x, y \in \mathbb{N}$, $S_0 f(x) = S_0 f(y) = 1$ and by formula (5),*

$$\begin{aligned}\partial_t(S_t f)(x)|_{t=0} &= (L_u f|_{\mathbb{N}})(x) + V_u(x)(f(-1) - f(x)) = -V_u(x), \\ \partial_t(S_t f(x) - S_t f(y))|_{t=0} &= V_u(y) - V_u(x).\end{aligned}$$

The function f is non-decreasing on $\mathbb{N} \cup \{-1\}$ and a necessary condition for $S_t f$ to be non-decreasing for all $t \geq 0$ is, in the light of the preceding equation, that $V_u(y) - V_u(x) \leq 0$ whenever $x \leq y$, i.e. V_u is non-increasing on \mathbb{N} .

If V_u is constant, then Theorem 2.2 admits a variant involving the gradient $\partial_v \partial_u$ instead of $\partial_v^* \partial_u$, which is stated in Theorem 2.5 below for the sake of completeness. In the applications, when V_u is constant, we choose to invoke Theorem 2.5 in lieu of Theorem 2.2, because the underlying arguments are much simpler. Indeed, in this case the equation (7) reduces to

$$\partial_u P_t f = e^{-V_u t} P_{u,t} \partial_u f, \quad t \geq 0,$$

and it is no longer required to extend artificially the Markov process, nor to add information at the boundary, in order to differentiate a second time. As a matter of fact, one can notice that if V_u is constant, then the BDP associated to the semigroup $(P_{u,*v,t})_{t \geq 0}$ of Theorem 2.2 do not visit the state 0 unless it starts there.

In order to state the theorem, a new birth-death semigroup $(P_{u,v,t})_{t \geq 0}$ with generator $L_{u,v}$ and a potential $V_{u,v}$ is introduced. Set for all real-valued functions on \mathbb{N} :

$$\begin{aligned}L_{u,v} f(x) &= \alpha_{u,v} \partial f(x) + \beta_{u,v} \partial^* f(x), & x \in \mathbb{N}, \\ \alpha_{u,v}(x) &= \frac{v(x+1)}{v(x)} \frac{u(x+2)}{u(x+1)} \alpha(x+2), \quad \beta_{u,v}(x) = \frac{v(x-1)}{v(x)} \frac{u(x-1)}{u(x)} \beta(x), & x \in \mathbb{N}, \\ V_{u,v}(x) &= \alpha(x) - \frac{v(x+1)}{v(x)} \frac{u_{x+2}}{u(x+1)} \alpha(x+2) + \left(\frac{u(x)}{u(x+1)} + 1 \right) \beta(x+1) \\ &\quad - \left(1 + \frac{v(x-1)}{v(x)} \right) \frac{u(x-1)}{u(x)} \beta(x), & x \in \mathbb{N}.\end{aligned}$$

In contrast to the Markov semigroup $(P_{u,*v,t})_{t \geq 0}$, the semigroup $(P_{u,v,t})_{t \geq 0}$ is always a birth-death semigroup.

Theorem 2.5 (Alternative version of the second intertwining relation). *Assume that V_u is constant on \mathbb{N} and that $V_{u,v}$ is bounded from below. For all real-valued functions on \mathbb{N} such that $\|\partial_u f\|_{\infty} < +\infty$ and $\|\partial_v \partial_u f\|_{\infty} < +\infty$, we have:*

$$\partial_v \partial_u (P_t f) = P_{u,v,t}^{V_{u,v}} (\partial_v \partial_u f), \quad t \geq 0.$$

Remark 2.6 (Link between the two versions of the second intertwining). *Surprisingly, it is only possible to deduce directly Theorem 2.5 from Theorem 2.2 in the case where the sequence v is constant. When $v = 1$ for instance, one can write that $\partial^* \partial_u f(\cdot + 1) = -\partial \partial_u f$, yielding under the appropriate assumptions on $f : \mathbb{N} \rightarrow \mathbb{R}$ that:*

$$P_{u,1,t} f(x) = P_{u,*1,t} \overleftarrow{f}(x+1), \quad x \in \mathbb{N}, \quad t \geq 0.$$

At the level of the processes, this equation can be reformulated into the equality in law:

$$X_{1,u,t}^x = X_{1,*u,t}^{x+1} - 1, \quad x \in \mathbb{N}, \quad t \geq 0,$$

where $(X_{1,*u,t}^x)_{t \geq 0}$ and $(X_{1,u,t}^x)_{t \geq 0}$ are the Markov processes corresponding respectively to the semigroups $(P_{u,*1,t})_{t \geq 0}$ and $(P_{u,1,t})_{t \geq 0}$. If v is not constant, no similar relation holds in general.

Remark 2.7 (Other versions). *It is possible to derive similar theorems for other gradients. For example, if the gradient ∂^* is defined as $\partial^* f = \partial^* f$ on \mathbb{N}^* and with the Neumann-like boundary condition in 0, $\partial^* f(0) = 0$, then the analogous theorem to Theorem 2.2 holds for $\partial_v \partial_u^*$. It is also possible to derive intertwining relations in the case where the semigroup lives on $\llbracket 0, n \rrbracket$, although the underlying structures are rather different: for instance, the condition V_u non-increasing is no longer necessary.*

Let us turn to our first application of Theorem 2.2 and its variant Theorem 2.5. The first order intertwining relation recalled in Theorem 2.1 yields a contraction property in Wasserstein distance. Precisely, under the assumptions of Theorem 2.1, by [Chafaï and Joulin, 2013, Corollary 3.1], we have for all $\mu, \nu \in \mathcal{P}$,

$$W_{d_u}(\mu P_t, \nu P_t) \leq e^{-\sigma(u)t} W_{d_u}(\mu, \nu), \quad (8)$$

where the distance d_u on \mathbb{N} and the related Wasserstein distance W_{d_u} on \mathcal{P} are defined in the forthcoming section, Section 3.1. Similarly, Theorems 2.2 and 2.5 lead to a contraction property for the distances $\zeta_{u,*v}$ and $\zeta_{u,v}$, defined respectively for two sequence of positive weights u and v by

$$\begin{aligned} \zeta_{u,*v} &= \sup_{f \in \mathcal{F}_{u,*v}} |\mu(f) - \nu(f)|, & \mathcal{F}_{u,*v} &= \{f : \mathbb{N} \rightarrow \mathbb{R}, \|\partial_u f\|_\infty < \infty, \|\partial_v^* \partial_u f\|_\infty \leq 1\}, \\ \zeta_{u,v} &= \sup_{f \in \mathcal{F}_{u,v}} |\mu(f) - \nu(f)|, & \mathcal{F}_{u,v} &= \{f : \mathbb{N} \rightarrow \mathbb{R}, \|\partial_u f\|_\infty < \infty, \|\partial_v \partial_u f\|_\infty \leq 1\}. \end{aligned}$$

We call $\zeta_{u,*v}$ and $\zeta_{u,v}$ second order Zolotarev-type distances since they are simple metric distances in the sense of Zolotarev (Zolotarev [1976]) and can be seen as the discrete counterparts of the distance ζ_2 defined on the set of real probability distributions (the distance ζ_2 , introduced in Zolotarev [1976] and further studied in Rio [1998], is associated to the set of continuously differentiable functions on \mathbb{R} whose derivative is Lipschitz). The contraction property reads as follows:

Theorem 2.8 (Contraction of the BDP in second order distances).

- Under the same hypotheses as in Theorem 2.2, we set $\sigma(u, *v) = \inf V_{u,*v}$. Then, for all $\mu, \nu \in \mathcal{P}$, we have:

$$\zeta_{u,*v}(\mu P_t, \nu P_t) \leq e^{-\sigma(u,*v)t} \zeta_{u,*v}(\mu, \nu). \quad (9)$$

- Under the assumptions of Theorem 2.5, define $\sigma(u, v) = \inf V_{u,v}$. Letting $\mu, \nu \in \mathcal{P}$, it stands that:

$$\zeta_{u,v}(\mu P_t, \nu P_t) \leq e^{-\sigma(u,v)t} \zeta_{u,v}(\mu, \nu). \quad (10)$$

Proof. The proof is done in the first case, the second one being similar. For all real-valued functions f on \mathbb{N} such that $\|\partial_u f\|_\infty < \infty$ and $\|\partial_v^* \partial_u f\|_\infty \leq 1$, Theorem 2.2 implies that

$$\|\partial_v^* \partial_u P_t f\|_\infty \leq e^{-\sigma(u, *v)t} \|\partial_v^* \partial_u f\|_\infty \leq e^{-\sigma(u, *v)t}, \quad t \geq 0.$$

Hence,

$$\begin{aligned} \zeta_{u, *v}(\mu P_t, \nu P_t) &= \sup \left\{ \left| \int P_t f d\mu - \int P_t f d\nu \right|, \|\partial_v^* \partial_u f\|_\infty \leq 1 \right\} \\ &\leq \sup \left\{ \left| \int g d\mu - \int g d\nu \right|, \|g\|_\infty \leq e^{-\sigma(u, *v)t} \right\} \\ &= e^{-\sigma(u, *v)t} \zeta_{u, *v}(\mu, \nu). \end{aligned}$$

□

If the quantity $\sigma(u, *v)$ (resp. $\sigma(u, v)$) is positive, the first bound (resp. the second) is a contraction. In particular, if we take $\nu = \pi$ the invariant measure of the BDP, then Theorem 2.8 gives the rate of convergence of the BDP towards its invariant measure in a second order distance.

Remark 2.9 (Generalization and optimality).

- The proof of Theorem 2.8 can be generalized to the Zolotarev-type distance associated to the set of functions $f : \mathbb{N} \rightarrow \mathbb{R}$ such that $\|Df\|_\infty \leq 1$ as soon as we have an inequality of the type $\|DP_t f\|_\infty \leq e^{-\sigma t} \|Df\|_\infty$ for every $t \geq 0$, some $\sigma > 0$ and some finite difference operator D . In Section 4 below, we detail such convergences in higher order Zolotarev-type distances.
- By arguments similar to those developed in [Chafaï and Joulin, 2013, corollary 3.1], one can prove that the constants $\sigma(u, *v)$ and $\sigma(u, v)$ in the equations (9) and (10) are optimal. Indeed, the argument of Chafaï and Joulin [2013] relies on the propagation of the monotonicity and we have the analogous property at the second order (cf Remark 2.3).
- Using [Chen, 2004, Theorem 9.25], we see that, choosing a good sequence u , it is possible to obtain the contraction in the Wasserstein distance (8) at a rate corresponding to the spectral gap (even if there is no corresponding eigenvector). For the second order, we do not know if it is always possible to find sequences u, v such that $\sigma(*u, v)$ or $\sigma(u, v)$ is equal to the second smallest positive eigenvalue of $-L$.

In the following section we focus our attention on our main application of intertwining relations, Stein's factors.

3 Application to Stein's magic factors

3.1 Distances between probability distributions

First of all, we introduce the distances between probability measures used to measure approximations in the sequel. They are of the form

$$\zeta_{\mathcal{F}}(\mu, \nu) = \sup \{ |\mu(f) - \nu(f)|, f \in \mathcal{F} \},$$

where \mathcal{F} is a subset of the set of real-valued functions on \mathbb{N} . The distances $\zeta_{u, *v}$, $\zeta_{u, v}$ presented at the end of the preceding section were examples of such distances; we now recall the definition of three classical distances on \mathcal{P} .

Total variation distance. The total variation distance d_{TV} is the distance associated to the set \mathcal{F}_{TV} of real-valued functions on \mathbb{N} such that $0 \leq f \leq 1$. In contrast to the continuous space case, the topology induced by the total variation distance on \mathbb{N} is exactly the convergence in law. Some authors prefer to define the total variation distance as the distance associated to the set $\mathcal{F} = \{f : \mathbb{N} \rightarrow \mathbb{R}, \|f\|_\infty \leq 1\}$. The two definitions vary by a factor $\frac{1}{2}$:

$$d_{\text{TV}}(\mu, \nu) = \sup_{0 \leq f \leq 1} |\mu(f) - \nu(f)| = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} |\mu(f) - \nu(f)| = \frac{1}{2} \sum_{x \in \mathbb{N}} |\mu(x) - \nu(x)|.$$

Wasserstein distance. For a distance d on \mathbb{N} let us call $\text{Lip}(d)$ the set of real-valued functions on \mathbb{N} such that

$$|f(x) - f(y)| \leq d(x, y), \quad x, y \in \mathbb{N}.$$

The Wasserstein distance between two probability measures μ and ν of \mathcal{P} is defined as

$$W_d(\mu, \nu) = \inf \int d(x, y) d\Pi(x, y),$$

where the infimum is taken over all probability measures Π on \mathbb{N}^2 whose first marginal is μ and second marginal is ν . By Kantorovich-Rubinstein theorem (see e.g. Szulga [1982]),

$$W_d(\mu, \nu) = \zeta_{\text{Lip}(d)}(\mu, \nu).$$

For a positive sequence u , define the distance d_u on \mathbb{N} as

$$d_u(x, y) = \sum_{k=x}^{y-1} u(k), \quad x < y; \quad d_u(x, y) = d_u(y, x), \quad x > y; \quad d_u(x, y) = 0, \quad x = y.$$

Let us observe that $\text{Lip}(d_u) = \{f : \mathbb{N} \rightarrow \mathbb{R}, \|\partial_u f\|_\infty \leq 1\}$. Hence

$$W_{d_u}(\mu, \nu) = \sup_{f \in \text{Lip}(d_u)} |\mu(f) - \nu(f)| = \sup_{\|\partial_u f\|_\infty \leq 1} |\mu(f) - \nu(f)|.$$

The distance associated to the constant sequence equal to 1 is the usual distance $d_1(x, y) = |x - y|$. We denote by $W = W_{d_1}$ the associated Wasserstein distance.

Kolmogorov distance. The Kolmogorov distance is defined as the metric distance associated to the set \mathcal{F}_K of indicator functions of intervals $[0, x]$:

$$d_K(\mu, \nu) = \sup_{x \in \mathbb{N}} |\mu([0, x]) - \nu([0, x])|.$$

Comparison between distances. For all $\mu, \nu \in \mathcal{P}$,

$$d_K(\mu, \nu) \leq d_{\text{TV}}(\mu, \nu) \leq \frac{1}{\inf_{\mathbb{N}} u} W_{d_u}(\mu, \nu).$$

Indeed, both inequalities are consequences of the inclusions

$$\mathcal{F}_K \subset \mathcal{F}_{\text{TV}} \subset \frac{1}{\inf_{\mathbb{N}} u} \text{Lip}(d_u).$$

The second inclusion follows from the implication

$$0 \leq f \leq 1 \Rightarrow \|\partial f\|_\infty \leq \frac{1}{\inf_{\mathbb{N}} u}.$$

The total variation distance is invariant by translation, whereas intuitively the Wasserstein distance gives more weight to the discrepancy between $\mu(x), \nu(x)$ if it occurs for a large integer x . The Kolmogorov distance may be used as an alternative to the total variation distance when the latter is too strong to measure the involved quantities.

3.2 Basic facts on Stein's method

Given a probability measure μ and a target probability measure π of \mathcal{P} , the Stein-Chen method provides a way to estimate the distances of the type $\zeta_{\mathcal{F}}(\mu, \pi)$. More precisely, consider a Stein's operator S :

$$Sf(x) = \alpha(x)f(x+1) - \beta(x)f(x), \quad x \in \mathbb{N}; \quad \beta_0 = 0,$$

characterizing the probability measure π (meaning that $\int Sf d\mu = 0$ for every function $f : \mathbb{N} \rightarrow \mathbb{R}$ in a sufficiently rich class of functions if and only if $\mu = \pi$) and the associated Stein equation

$$Sg_f = f - \int f d\pi. \quad (11)$$

We call g_f a solution to Stein's equation. The interest of such solutions comes from the following error bound:

$$\zeta_{\mathcal{F}}(\mu, \pi) = \sup_{f \in \mathcal{F}} |\mu(f) - \pi(f)| = \sup_{f \in \mathcal{F}} \left| \int Sg_f d\mu \right|. \quad (12)$$

As a consequence, if it can be shown that

$$\left| \int Sg_f d\mu \right| \leq \varepsilon_0 \|g_f\|_{\infty} + \varepsilon_1 \|\partial g_f\|_{\infty},$$

then it follows that

$$\zeta_{\mathcal{F}}(\mu, \pi) \leq \varepsilon_0 \sup_{f \in \mathcal{F}} \|g_f\|_{\infty} + \varepsilon_1 \sup_{f \in \mathcal{F}} \|\partial g_f\|_{\infty}.$$

This strategy of proof is widely used, for example in the references about Stein's method provided in the introduction.

A key point of this approach consists then in evaluating the so-called first and second Stein factors, also known as *magic factors*:

$$\sup_{f \in \mathcal{F}} \|g_f\|_{\infty}, \quad \sup_{f \in \mathcal{F}} \|\partial g_f\|_{\infty}.$$

Observe that the equation (11) does not determine the value of $g_f(0)$. When evaluating the first Stein factor $\sup_{f \in \mathcal{F}} \|g_f\|_{\infty}$, we pick for every $f \in \mathcal{F}$ the solution g_f of (11) such that $g_f(0) = 0$. Hence, it is sufficient to consider the quantity

$$\sup_{f \in \mathcal{F}} \|g_f\|_{\infty, \mathbb{N}^*} = \sup_{f \in \mathcal{F}} \|\vec{g_f}\|_{\infty}.$$

Similarly, for the second Stein factor, picking solutions g_f to (11) satisfying to $g_f(0) = g_f(1)$, i.e. $\partial g_f(0) = 0$, allows to consider only the quantity

$$\sup_{f \in \mathcal{F}} \|\partial g_f\|_{\infty, \mathbb{N}^*} = \sup_{f \in \mathcal{F}} \|\vec{\partial g_f}\|_{\infty}.$$

To evaluate the above quantities, we use a method known as method of the generator and the semigroup representation deriving from it. Set L the generator and $(P_t)_{t \geq 0}$ the semigroup associated to the BDP(α, β) and assume that $(P_t)_{t \geq 0}$ is invariant with respect to the target probability distribution π . The operators S and L are linked by the relation

$$Lh = S(-\partial^* h).$$

The Poisson equation reads as

$$Lh_f = f - \mu(f),$$

the centered solution h_f being given by the expression

$$h_f = - \int_0^\infty (P_t f - \mu(f)) dt.$$

Then, we obtain a solution g_f to Stein's equation (11) under the so-called semigroup representation:

$$g_f = -\partial^* h_f = \int_0^\infty \partial^* P_t f dt. \quad (13)$$

3.3 Bounds on Stein's magic factors

In this section, theoretical bounds on the first and second order Stein factors are proposed for the approximation in total variation, Wasserstein and Kolmogorov distances. Proofs are postponed to Section 6 in order to clarify the presentation. Before turning to the results, a few general comments are made.

1. Our method evaluates Stein factors by quantities of the form

$$\int_0^\infty e^{-\kappa t} \sup_{i \in \mathbb{N}} \mathbb{P}(\tilde{X}_t^i = i) dt, \quad \int_0^\infty e^{-\kappa t} \sup_{i \in \mathbb{N}^*} \left(\mathbb{P}(\tilde{X}_t^i = i) - \mathbb{P}(\tilde{X}_t^i = i-1) \right) dt.$$

The Markov process $(\tilde{X}_t)_{t \geq 0}$ which occurs is an alternative process and is not necessarily the same as the BDP (α, β) with semigroup $(P_t)_{t \geq 0}$ appearing in the semigroup representation (13). To our knowledge, this is new and makes the originality of our work.

2. While the detailed demonstrations of the forthcoming results are given in Section 6, the scheme of proof is briefly explained here. Firstly, the argmax f_i of the pointwise Stein factor

$$\sup_{f \in \mathcal{F}} |\partial^k g_f(i)|, \quad i \in \mathbb{N}^*, \quad k \in \{0, 1\},$$

is obtained by resuming and generalizing results from Brown and Xia [2001]. Secondly, the function f_i is plugged in the semigroup representation:

$$\partial^k g_{f_i}(i) = \int_0^\infty \partial^k \partial^* P_t f_i dt, \quad k \in \{0, 1\}.$$

The intertwining relations of Section 2 are then used to rewrite the term $\partial^k \partial^* P_t$.

This technique is already employed for Poisson approximation in some works, Barbour and Brown [1992] and Barbour and Xia [2006] for example. In that context, the intertwining relation reads as:

$$\partial P_t = e^{-t} P_t \partial, \quad t \geq 0,$$

where the semigroup $(P_t)_{t \geq 0}$ is the same on the left and on the right. The use of the intertwining relations permits to go beyond this case and to construct a universal method to derive Stein's factors.

3. For the sake of clarity, the present section only includes results on the uniform Stein factors. However, it can be seen in Section 6 that our upper bounds on the pointwise Stein factors are often sharp.

4. For the second order Stein factor, two sets of assumptions are used:

Assumptions 3.1 (Assumptions).

H₁: The potential V_1 is non-increasing and non-negative, the potential $V_{1,*u}$ is bounded from below, and the sequence u is bounded from below by a positive constant. In this case, we define $\sigma(1, *u) = \inf_{\mathbb{N}} V_{1,*u}$ and denote by $(X_{1,*u,t}^i)_{t \geq 0}$ the Markov process of generator $L_{1,*u}$ such that $X_{1,*u,0}^i = i$.

H₂: The potential V_1 is a non-negative constant and the potential $V_{1,u}$ is bounded from below. In this case, set $\sigma(1, u) = \inf_{\mathbb{N}} V_{1,u}$ and call $(X_{1,u,t}^i)_{t \geq 0}$ the birth-death process of generator $L_{1,u}$ such that $X_{1,u,0}^i = i$.

This comes from the fact that the double intertwining relation is given by the main result Theorem 2.2 under **H₁** and by its analogous Theorem 2.5 under **H₂**.

5. Stein's factors related to the different distances compare between each other through the inequalities:

$$\sup_{f=\mathbf{1}_{[0,m]}, m \in \mathbb{N}} \|\partial^k g_f\|_{\infty} \leq \sup_{0 \leq f \leq 1} \|\partial^k g_f\|_{\infty} \leq \frac{1}{\inf_{\mathbb{N}} u} \sup_{f \in \text{Lip}(d_u)} \|\partial^k g_f\|_{\infty}, \quad k \in \mathbb{N}.$$

We now state the main results of this section, formulated for each distance of interest.

Approximation in total variation distance.

Theorem 3.2 (First Stein's factor for bounded functions). Assume that V_u is bounded from below by some positive constant $\sigma(u)$. Then, we have:

$$\sup_{0 \leq f \leq 1} \|g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(u)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{u,t}^i = i) dt.$$

This theorem is applied to the negative binomial approximation in Proposition 4.5.

Theorem 3.3 (Second Stein's factor for bounded functions I). Under **H₁**,

$$\sup_{0 \leq f \leq 1} \|\partial g_f\|_{\infty} \leq 2 \int_0^{\infty} e^{-\sigma(1,*u)t} \sup_{i \in \mathbb{N}^*} \mathbb{P}(X_{1,*u,t}^i = i) dt.$$

If the sequence is chosen to be $u = 1$, we have:

$$\begin{aligned} \sup_{0 \leq f \leq 1} \|\partial g_f\|_{\infty} &\leq \int_0^{\infty} e^{-\sigma(1,*u)t} \sup_{i \in \mathbb{N}^*} (\mathbb{P}(X_{1,*u,t}^i = i) - \mathbb{P}(X_{1,*u,t}^i = i-1) \\ &\quad + \mathbb{P}(X_{1,*u,t}^i = i) - \mathbb{P}(X_{1,*u,t}^i = i+1)) dt. \end{aligned}$$

The analogue of Theorem 3.3 under the alternative set of hypotheses reads as:

Theorem 3.4 (Second Stein's factor for bounded functions II). Under **H₂**,

$$\sup_{0 \leq f \leq 1} \|\partial g_f\|_{\infty} \leq 2 \int_0^{\infty} e^{-\sigma(1,u)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{1,u,t}^i = i) dt.$$

If the sequence is chosen to be $u = 1$, we have:

$$\begin{aligned} \sup_{0 \leq f \leq 1} \|\partial g_f\|_{\infty} &\leq \int_0^{\infty} e^{-\sigma(1,u)t} \sup_{i \in \mathbb{N}} (\mathbb{P}(X_{1,u,t}^i = i) - \mathbb{P}(X_{1,u,t}^i = i-1) \\ &\quad + \mathbb{P}(X_{1,u,t}^i = i) - \mathbb{P}(X_{1,u,t}^i = i+1)) dt. \end{aligned}$$

Remark 3.5 (Alternative versions). *By the same techniques, it is possible to upper bound the quantities*

$$\sup_{0 \leq f/u \leq 1} \|g_f\|_\infty, \quad \sup_{0 \leq f/u \leq 1} \|\partial_u g_f\|_\infty.$$

It could be useful if one is interested in the approximation in V -norm (Meyn and Tweedie [1993]) rather than in total variation distance.

Approximation in Wasserstein distance.

Theorem 3.6 (First Stein's factor for Lipschitz functions). *If V_u is bounded from below by some positive constant $\sigma(u)$, then we have:*

$$\sup_{f \in \text{Lip}(d_u)} \|\vec{g}_f/u\|_\infty \leq \frac{1}{\sigma(u)}.$$

Moreover, if V_u is constant, then the preceding inequality is in fact an equality.

Theorem 3.7 (Second Stein's factor for Lipschitz functions I). *Under \mathbf{H}_1 ,*

$$\sup_{f \in \text{Lip}(d_u)} \|\partial_u g_f\|_\infty \leq \frac{1}{\sigma(1, *u)} \sup_{x \in \mathbb{N}^*} \left(1 + \frac{u(x-1)}{u(x)} \right).$$

If we assume that $u(x) = q^x$ on \mathbb{N} with $q \geq 1$, then it stands that:

$$\sup_{f \in \text{Lip}(d_u)} \|\partial_u g_f\|_\infty \leq \int_0^\infty e^{-\sigma(1, *u)t} \left(1 - \frac{1}{q} + 2\frac{1}{q} \sup_{i \in \mathbb{N}^*} \mathbb{P}(X_{1, *u, t}^i = i) \right) dt.$$

An instance of Theorems 3.6 and 3.7 in the context of geometric approximation is given by Proposition 4.7. The following theorem is the analogue of Theorem 3.7 under the alternative set of hypotheses.

Theorem 3.8 (Second Stein's factor for Lipschitz functions II). *Under \mathbf{H}_2 ,*

$$\sup_{f \in \text{Lip}(d_u)} \left\| \frac{1}{u} \vec{\partial} g_f \right\|_\infty \leq \frac{1}{\sigma(1, u)} \sup_{x \in \mathbb{N}} \left(1 + \frac{u(x+1)}{u(x)} \right).$$

If the sequence is chosen to be $u(x) = q^x$ on \mathbb{N} with $q \geq 1$, then the following result holds:

$$\sup_{f \in \text{Lip}(d_u)} \left\| \frac{1}{u} \vec{\partial} g_f \right\|_\infty \leq \int_0^\infty e^{-\sigma(1, u)t} \left(q - 1 + 2 \sup_{i \in \mathbb{N}} \mathbb{P}(X_{1, u, t}^i = i) \right) dt.$$

As an illustration of this theorem, we derive Proposition 4.4 in the case of negative binomial approximation.

Approximation in Kolmogorov distance.

The first theorem indicates that the inequality

$$\sup_{1_{[0, m]}, m \in \mathbb{N}} \|g_f\|_\infty \leq \sup_{0 \leq f \leq 1} \|g_f\|_\infty$$

is actually an equality. This comes from the fact that the function achieving the argmax of the pointwise factor for bounded functions is actually of the form $f = 1_{[0, m]}$. As a consequence, our upper bounds for the first Stein factor are identical for the approximation in total variation and Kolmogorov distances.

Theorem 3.9 (First Stein's factor for indicator functions). *If V_u is bounded from below and $\inf_{\mathbb{N}} V_u = \sigma(u)$, it stands that:*

$$\sup_{1_{[0,m]}, m \in \mathbb{N}} \|g_f\|_{\infty} = \sup_{0 \leq f \leq 1} \|g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(u)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{u,t}^i = i) dt.$$

The two following theorems deal with the second Stein factor under the two set of hypotheses.

Theorem 3.10 (Second Stein's factor for indicator functions I). *Under \mathbf{H}_1 ,*

$$\sup_{f=1_{[0,m]}, m \in \mathbb{N}} \|\partial g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(1,*u)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{1,*u,t}^i = i) dt.$$

If the sequence is chosen to be $u = 1$, we have:

$$\sup_{f=1_{[0,m]}, m \in \mathbb{N}} \|\partial g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(1,*u)t} \sup_{i \in \mathbb{N}^*} |\mathbb{P}(X_{1,*u,t}^i = i) - \mathbb{P}(X_{1,*u,t}^i = i-1)| dt.$$

Comparing this second bound with the second bound obtained in Theorem 3.3 for the total variation approximation, one notices the fact that the second Stein factor for the total variation approximation involves the second derivative of the function $f_i(x) = \mathbb{P}(X_{1,*u,t}^i = x)$, whereas the second Stein factor for the Kolmogorov approximation involves only its first derivative.

Finally, we state the analogous of Theorem 3.10 under the alternative set of hypotheses.

Theorem 3.11 (Second Stein's factor for indicator functions II). *Under \mathbf{H}_2 ,*

$$\sup_{f=1_{[0,m]}, m \in \mathbb{N}} \|\partial g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(1,u)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{1,u,t}^i = i) dt.$$

If the sequence is chosen to be $u = 1$, we have:

$$\sup_{0 \leq f \leq 1} \|\partial g_f\|_{\infty} \leq \int_0^{\infty} e^{-\sigma(1,u)t} \sup_{i \in \mathbb{N}} |\mathbb{P}(X_{1,u,t}^i = i) - \mathbb{P}(X_{1,u,t}^i = i-1)| dt.$$

3.4 Stein's method and mixture of distributions

As another part of our work within the context of Stein's method, we present in the current section theoretical error bounds for the approximation of mixture of distributions. This section is independent from our study of Stein's factors contained in Section 3.3. Results from both sections are combined in Section 4 and applied to Poisson and geometric mixture approximation. Let φ be a non-negative function on \mathbb{N} such that $\varphi(0) = 0$. For $\lambda > 0$, we denote by $\mathcal{I}_{\varphi}(\lambda)$ the probability distribution on \mathbb{N} whose Stein's operator is

$$S_{\lambda}g(x) = \lambda g(x+1) - \varphi(x)g(x), \quad x \in \mathbb{N}.$$

By letting φ vary, one finds back for $\mathcal{I}_{\varphi}(\lambda)$ every probability distribution supported on \mathbb{N} . In particular, the choice $\varphi(x) = x$ gives the Poisson law and is studied in Barbour et al. [1992]. The choice $\varphi(x) = r + x$ and $\varphi(x) = 1$ leads respectively to the binomial negative and geometric laws. A less classical example is $\varphi(x) = x^2$, for which $\mathcal{I}_{\varphi}(\lambda)$ is a distribution with pointwise probabilities $C_{\lambda} \lambda^x / (x!)^2$ for $x \in \mathbb{N}$ (C_{λ} is the renormalizing constant).

The first theorem of this section reads as follows.

Theorem 3.12 (Closeness of two $\mathcal{I}_{\varphi}(\lambda)$ distributions). *Set $\lambda, \lambda' > 0$. We have:*

$$d_{\mathcal{F}}(\mathcal{I}_{\varphi}(\lambda'), \mathcal{I}_{\varphi}(\lambda)) \leq |\lambda - \lambda'| \sup_{f \in \mathcal{F}} \|g_{\lambda,f}\|_{\infty},$$

where g_f is the solution of the Stein's equation $S_{\lambda}g_f = f - \int f d\mathcal{I}_{\varphi}(\lambda)$. More generally, for any positive sequence u , if $X \sim \mathcal{I}_{\varphi}(\lambda)$ and $X' \sim \mathcal{I}_{\varphi}(\lambda')$, then:

$$d_{\mathcal{F}}(\mathcal{I}_{\varphi}(\lambda'), \mathcal{I}_{\varphi}(\lambda)) \leq |\lambda - \lambda'| \sup_{f \in \mathcal{F}} \|g_f/u\|_{\infty} \mathbb{E}[u(X' + 1)].$$

Proof. By the usual Stein error bound (12),

$$d_{\mathcal{F}}(\mathcal{I}_{\varphi}(\lambda'), \mathcal{I}_{\varphi}(\lambda)) = \sup_{f \in \mathcal{F}} \left| \int f d\mathcal{I}_{\varphi}(\lambda') - \int f d\mathcal{I}_{\varphi}(\lambda) \right| = \sup_{f \in \mathcal{F}} |\mathbb{E}[S_{\lambda} g_f(X')]|,$$

where $X' \sim \mathcal{I}_{\varphi}(\lambda')$. We know that $\mathbb{E}[S_{\lambda'} g_f(X')] = 0$; this yields:

$$\begin{aligned} |\mathbb{E}[S_{\lambda} g_f(X')]| &= |\mathbb{E}[(\lambda - \lambda') g_f(X' + 1)]| = |(\lambda - \lambda') \mathbb{E}[u(X' + 1) g_f(X' + 1) / u(X' + 1)]| \\ &\leq |\lambda - \lambda'| \|g_f / u\|_{\infty} \mathbb{E}[u(X' + 1)]. \end{aligned}$$

□

Note that the right hand side of both inequalities stated in Theorem 3.12 is not symmetric in (λ, λ') due to the dependence of g_f on λ and one can slightly improve it by taking the minimum over the symmetrized form. The first inequality corresponds to the constant sequence $u = 1$. Let W be a mixture of law $\mathcal{I}_{\varphi}(\lambda)$; namely there exists a random variable Λ on \mathbb{R}^+ such that

$$\mathcal{L}(W \mid \Lambda) = \mathcal{I}_{\varphi}(\Lambda).$$

(Recall that $\mathcal{L}(W)$ denotes the distribution of the random variable W .) A consequence of Theorem 3.12 is the following corollary.

Corollary 3.13 (Biased approximation of mixed $\mathcal{I}_{\varphi}(\lambda)$ laws). *With the preceding notation, we have:*

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_{\varphi}(\lambda)) \leq \mathbb{E}[|\lambda - \Lambda|] \sup_{f \in \mathcal{F}} \|g_f\|_{\infty}.$$

Proof. Indeed,

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_{\varphi}(\lambda)) \leq \mathbb{E}[d_{\mathcal{F}}(\mathcal{L}(W \mid \Lambda), \mathcal{I}_{\varphi}(\lambda))] \leq \mathbb{E}[|\lambda - \Lambda|] \sup_{f \in \mathcal{F}} \|g_f\|_{\infty}.$$

□

However, one actually has the following better bound using the mixture property of W :

Theorem 3.14 (Unbiased approximation of mixed $\mathcal{I}_{\varphi}(\lambda)$ distributions). *For every positive sequence u , letting $\lambda = E[\Lambda]$, we have:*

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_{\varphi}(\lambda)) \leq \sup_{f \in \mathcal{F}} \|\partial_u g_f\|_{\infty} \sup_{f \in \text{Lip}(d_{\vec{u}})} \|g_f\|_{\infty} \text{Var}(\Lambda).$$

More generally, the following upper bound holds for all positive sequences u, v :

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{I}_{\varphi}(\lambda)) \leq \sup_{f \in \mathcal{F}} \|\partial_u g_f\|_{\infty} \sup_{r \in \text{Lip}(d_{\vec{u}})} \|g_r / v\|_{\infty} \mathbb{E}[|\lambda - \Lambda|^2 \mathbb{E}[v(W + 1) \mid \Lambda]].$$

Proof of Theorem 3.14. For every real-valued function g on \mathbb{N} , $\mathbb{E}[S_{\Lambda}(g)(W) \mid \Lambda] = 0$. Hence, by taking $g = g_f$ the solution to Stein's equation associated with any fixed function $f : \mathbb{N} \rightarrow \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[S_{\Lambda} g_f(W)] &= \mathbb{E}[\mathbb{E}[(S_{\Lambda} - S_{\Lambda}) g_f(W) \mid \Lambda]] = \mathbb{E}[(\lambda - \Lambda) g_f(W + 1)] \\ &= \mathbb{E}[(\lambda - \Lambda)(g_f(W + 1) - g_f(Z + 1))], \end{aligned}$$

where $Z \sim \mathcal{I}_{\varphi}(\lambda)$. For two random variables Z, Z' on \mathbb{N} , by the Kantorovich-Rubinstein theorem recalled at the beginning of this section,

$$\begin{aligned} |\mathbb{E}[g(Z' + 1) - g(Z + 1)]| &\leq \|\partial_u g\|_{\infty} W_{d_u}(\mathcal{L}(Z' + 1), \mathcal{L}(Z + 1)) = \|\partial_u g\|_{\infty} \inf \mathbb{E}[d_u(Z' + 1, Z + 1)] \\ &= \|\partial_u g\|_{\infty} W_{d_{\vec{u}}}(\mathcal{L}(Z'), \mathcal{L}(Z)), \end{aligned}$$

where the infimum is taken on the set of couplings with first marginal $\mathcal{L}(Z)$ and second marginal $\mathcal{L}(Z')$. Now, by Theorem 3.12,

$$\begin{aligned}\mathbb{E}[S_\lambda g_f(W)] &\leq \|\partial_u g\|_\infty \mathbb{E}[|\lambda - \Lambda| \mathbb{E}[W_{d_u}(\mathcal{L}(W), \mathcal{L}(Z)) | \Lambda]] \\ &\leq \|\partial_u g\|_\infty \sup_{r \in \text{Lip}(d_u)} \|g_r/v\|_\infty \mathbb{E}[(\lambda - \Lambda)^2 \mathbb{E}[v(W+1) | \Lambda]].\end{aligned}$$

As in Theorem 3.12, the first inequality is an instance of the second one in the case $v = 1$. \square

Remark 3.15 (Alternative bound via coupling). *In the previous proof, we used Theorem 3.12 in order to bound $W_{d_u}(\mathcal{I}_\varphi(\Lambda), \mathcal{I}_\varphi(\lambda))$. It is also possible to bound this distance via another method (for instance a coupling argument) instead of using a bound on Stein's solution.*

4 Examples

In this section we illustrate our results on some examples. The classical examples of the M/M/ ∞ and M/M/1 process come from the queueing theory. We also apply the results to the Galton-Watson process with immigration. Other explicit examples of birth-death processes for which a "good choice" of sequence u is known are given in [Chen, 2004, Table 9.1 p. 351] and in Chen [1996]. For the sake of conciseness we defer the proof of Lemma 7.1 about the pointwise probabilities of the M/M/ ∞ queue to Section 7.

4.1 The M/M/ ∞ process and the Poisson approximation

Let $(X_t)_{t \geq 0}$ be a BDP with constant birth death λ and linear death rate $x \mapsto x$. Its invariant measure is the Poisson law \mathcal{P}_λ . Let us set $u = v = 1$ on \mathbb{N} . By application of Theorem 2.1 we find that $V_1 = 1$ and that $(P_{1,t})_{t \geq 0} = (P_t)_{t \geq 0}$. Applying Theorem 2.5 (or re-applying Theorem 2.1) yields $V_{1,1} = 2$ and $(P_{1,1,t})_{t \geq 0} = (P_t)_{t \geq 0}$. By a straightforward induction, for all positive or bounded functions $f : \mathbb{N} \rightarrow \mathbb{R}$,

$$\partial^k P_t f = e^{-kt} P_t \partial^k f, \quad t \geq 0, \quad k \in \mathbb{N}. \quad (14)$$

Combined with Theorem 2.8 and Remark 2.9, the equation (14) implies the following contraction in Zolotarev-type distance: for all $\mu \in \mathcal{P}$,

$$\sup_{\|\partial^k f\|_\infty \leq 1} |\mu(P_t f) - \mathcal{P}_\lambda(f)| \leq e^{-kt} \sup_{\|\partial^k f\|_\infty \leq 1} |\mu(f) - \mathcal{P}_\lambda(f)|, \quad k \in \mathbb{N}^*.$$

Formula (14) is already known and often proved using Mehler's formula which reads, for any bounded function f , as:

$$P_t f(x) = \mathbb{E}[f(X_t^0 + B_t)], \quad x \in \mathbb{N}, \quad t \geq 0, \quad (15)$$

where $(X_t^0)_{t \geq 0}$ is a M/M/ ∞ process starting from 0 and B_t is an independent random variable distributed as a binomial random variable with parameters (x, e^{-t}) . It is also known that X_t^0 is distributed as a Poisson distribution with parameter $\lambda(1 - e^{-t})$ at every time $t \geq 0$. Conversely, the proof of the formula (15) can be deduced from Theorem 2.1 with similar (but simpler) arguments than those developed in Lemma 4.2 below.

We now turn to the subject of Poisson approximation and the associated Stein factors. Let g_f be the solution to Stein's equation (11) with Stein's operator $Sf(x) = \lambda f(x+1) - xf(x)$. The target measure is the Poisson distribution \mathcal{P}_λ . The following lemma allows to estimate from above the pointwise probabilities of the process $(X_t)_{t \geq 0}$.

Lemma 4.1 (Upper bounds of the instantaneous probabilities of the M/M/ ∞ queue). *Let $(X_t^x)_{t \geq 0}$ be a BDP(λ, x) $_{x \in \mathbb{N}}$. For all $x \in \mathbb{N}$ and $t \geq 0$,*

$$\sup_{x \in \mathbb{N}} \mathbb{P}(X_t^x = x) \leq 1 \wedge \frac{c}{\sqrt{\lambda(1 - e^{-t})}}, \quad c = \frac{1}{\sqrt{2e}},$$

$$\sup_{x \in \mathbb{N}^*} |\mathbb{P}(X_t^x = x) - \mathbb{P}(X_t^x = x - 1)| \leq 1 \wedge \frac{C}{\lambda(1 - e^{-t})}, \quad C = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \leq 1.$$

The first upper bound is very classical, it derives from Mehler's formula (15) and an upper bound on the pointwise probabilities of the Poisson distribution ([Barbour et al., 1992, Proposition A.2.7]). The second one is new and is proved in Section 7, since it is rather technical and can be omitted at first reading.

By applying Theorems 3.2, 3.6, 3.8 jointly with the first bound of Lemma 4.1, one finds back (and by the same techniques) the following upper bounds (Barbour and Brown [1992], Barbour and Xia [2006]):

$$\sup_{0 \leq f \leq 1} \|g_f\|_\infty \leq 1 \wedge \sqrt{\frac{2}{\lambda e}}, \quad \sup_{f \in \text{Lip}(d_1)} \|g_f\|_\infty = 1, \quad \sup_{f \in \text{Lip}(d_1)} \|\partial g_f\|_\infty \leq 1 \wedge \frac{8}{3\sqrt{2e\lambda}}.$$

Of course, one may want to derive other known Stein's factors for Poisson approximation by our techniques, as for instance the second Stein factor for approximation in the total variation distance with rate $1 \wedge (1/\lambda)$ (Barbour et al. [1992]). However, when applying Theorem 3.4 with the second bound of Lemma 4.1, the non-integrability in 0 of the term $1/(1 - e^{-t})$ leads to sub-optimal results (namely, after some careful computations, we recover the known rate, up to a multiplicative factor $\log \lambda$).

Let us now combine the Stein bounds with our results on the mixture of distributions. If $\varphi(x) = x$ then $\mathcal{I}_\varphi(\lambda) = \mathcal{P}_\lambda$. In particular, Theorem 3.12 and the preceding bounds give

$$d_{\text{TV}}(\mathcal{P}(\lambda), \mathcal{P}(\lambda')) \leq \frac{1}{1 \wedge \sqrt{\lambda \vee \lambda'}} |\lambda - \lambda'|, \quad W(\mathcal{P}_\lambda, \mathcal{P}_{\lambda'}) \leq |\lambda - \lambda'|.$$

The first bound is (almost) the result of [Barbour et al., 1992, Theorem 1.C p. 12]. The second one is in fact an equality and can also be proved via a coupling approach (Lindvall [2002]). Theorem 3.14 yields

$$W(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \left(1 \wedge \frac{8}{3\sqrt{2e\lambda}}\right) \text{Var}(\Lambda), \quad d_{\text{TV}}(\mathcal{L}(W), \mathcal{I}_\varphi(\lambda)) \leq \frac{1}{\lambda} \text{Var}(\Lambda).$$

While the second bound is exactly the same as in [Barbour et al., 1992, Theorem 1.C p. 12], the bound in Wasserstein distance seems to be new. Let us see an instance of it. We denote by $\text{NB}(r, p)$ the negative binomial distribution of parameters (r, p) , i.e.,

$$\text{NB}(r, p)(x) = \frac{\Gamma(r+x)}{\Gamma(r)x!} (1-p)^r p^x, \quad x \in \mathbb{N},$$

where Γ denotes the usual Γ function. The negative binomial law is a mixed Poisson distribution with Λ distributed as a Gamma law with parameters r and $\frac{1-p}{p}$. Consequently, we obtain

$$W(\text{NB}(r, p), \mathcal{P}_{r(1-p)/p}) \leq \frac{8}{3\sqrt{2e}} \sqrt{\frac{r(1-p)}{p} \frac{(1-p)}{p}},$$

which is the upper bound announced in the introduction. A similar approximation in total variation distance holds. Although the convergence of the binomial negative distribution towards a Poisson law in the regime $p \rightarrow 1$, $r \rightarrow \infty$ and $r(1-p)/p \rightarrow c$ for a positive constant c is a well-known fact, the preceding upper bound seems to be the first attempt to quantify this convergence.

4.2 The GWI process and the negative binomial approximation

We consider the BDP with rates $\alpha(x) = p(r + x)$, $\beta(x) = x$ on \mathbb{N} with $r > 0$ and $0 < p < 1$. The coefficient pr can be interpreted as a rate of immigration, while the birth rate per capita is p and the death rate per capita is 1. Without the immigration procedure, this is a Galton-Watson process whose individuals have only one descendant (or simply a linear birth-death process). The invariant measure of this process is the negative binomial distribution $\text{NB}(r, p)$ just defined. Remark that for the particular choice $r = 1$ it is nothing else than the geometric law of parameter p . If X is a $\text{NB}(r, p)$ random variable then $X + r$ follows the so-called Pascal distribution; it represents the number of successes in a sequence of independent and identically distributed Bernoulli trials (with parameter p) before r failures when r is a positive integer. Let us take $u = v = 1$ on \mathbb{N} . Theorem 2.1 shows that:

$$\partial P_t = P_{1,t}^{V_1}, \quad t \geq 0,$$

where $(P_{1,t})_{t \geq 0}$ is a birth-death process with rates defined as

$$\alpha_1(x) = p(r + 1 + x), \quad \beta_1(x) = x, \quad x \in \mathbb{N}.$$

It is again a Galton-Watson process with immigration. The birth and death rates are unchanged and the immigration rate is increased by p . The potential V_1 is constant and takes the value $V_1 = 1 - p$. By Theorem 2.5, we find that $(P_t)_{t \geq 0}$ and ∂^2 are intertwined via the Feynman-Kac semigroup composed of a birth-death semigroup with rates $(\alpha_{1,1}, \beta_{1,1})$ and of potential $V_{1,1}$, with:

$$\alpha_{1,1}(x) = p(r + 2 + x), \quad \beta_{1,1}(x) = x, \quad V_{1,1}(x) = 2(1 - p), \quad x \in \mathbb{N}.$$

Let us call $(P_{k,t})_{t \geq 0}$ the semigroup associated to a BDP with rates $(p(r + k + x), x)$ on \mathbb{N} . By a straightforward induction, for all positive or bounded functions $f : \mathbb{N} \rightarrow \mathbb{R}$, the following intertwining relation holds:

$$\partial^k P_t f = e^{-(1-p)kt} P_{k,t} \partial^k f, \quad t \geq 0, \quad k \in \mathbb{N}. \quad (16)$$

As indicated in Remark 2.9, the previous equality gives the following improvement of Theorem 2.8: for every $\mu \in \mathcal{P}$,

$$\sup_{\|\partial^k f\|_\infty \leq 1} |\mu(P_t f) - \pi(f)| \leq e^{-(1-p)kt} \sup_{\|\partial^k f\|_\infty \leq 1} |\mu(f) - \pi(f)|.$$

Another consequence of the formula (16) is the invariance of polynomials under the action of $(P_t)_{t \geq 0}$: if Q is a polynomial of degree k , then for all $t \geq 0$, $\partial^k P_t Q$ is constant, hence $P_t Q$ is still a polynomial of degree k . This property also holds for the M/M/ ∞ process.

Intertwining relations can be seen in certain cases as consequences of Mehler-type formulas. Here, conversely, we are able to derive a Mehler-type formula from the first order intertwining relation. To our knowledge, this formula is new, though another Mehler-type formula is proved in Barbour et al. [2015].

Lemma 4.2 (A Mehler's formula for the Galton-Watson process with immigration). *Set $0 < p < 1$, $s > 0$ and $q = 1 - p$. For all $x \in \mathbb{N}$ let $(Y_t^x)_{t \geq 0}$ be a birth-death process starting at x and with rates $(p(s + k), k)_{k \in \mathbb{N}}$. Let W_t be a random variable following the Poisson distribution $\mathcal{P}(p(1 - e^{-t}))$ and define the sequence $(w(k))_{k \in \mathbb{N}}$ as*

$$w(0) = 1 - e^{-qt} \mathbb{P}(W_t = 0) \quad \text{and} \quad \forall k \in \mathbb{N}^*, \quad w(k) = e^{-qt} (\mathbb{P}(W_t = k - 1) - \mathbb{P}(W_t = k)).$$

For all $t \geq 0$, let the random variables $(Z_{i,t})_{i \in \mathbb{N}}$ be independent, identically distributed and independent of Y_t^0 , with distribution given by the pointwise probabilities $(w(k))_{k \in \mathbb{N}}$. Then we have the equality in law

$$Y_t^x = Y_t^0 + \sum_{i=1}^x Z_{i,t}.$$

Proof. This proof is a corollary of the intertwining formula (16) for $k = 1$. Indeed, for every bounded real-valued function on \mathbb{N} , Theorem 2.1 implies that

$$\mathbb{E}[f(Y_t^{x+1})] = \mathbb{E}[f(Y_t^x)] + e^{-qt} \mathbb{E}[f(\tilde{Y}_t^x + 1) - f(\tilde{Y}_t^x)], \quad x \in \mathbb{N}, \quad t \geq 0,$$

where $(\tilde{Y}_t^x)_{t \geq 0}$ is a BDP $(p(s+1+k), k)_{k \in \mathbb{N}}$. We notice that $(\tilde{Y}_t^x)_{t \geq 0} = (Y_t^x + W_t)_{t \geq 0}$, where $(W_t)_{t \geq 0}$ is a birth-death process independent of $(Y_t^x)_{t \geq 0}$ with rates $(p, k)_{k \in \mathbb{N}}$ such that $W_0 = 0$. The process $(W_t)_{t \geq 0}$ is a M/M/ ∞ queue starting from 0 at time 0. It is distributed as a Poisson law \mathcal{P}_{λ_t} , $\lambda_t = p(1 - e^{-t})$ at all times $t \geq 0$. We use below the observation that as $\lambda_t < 1$ for all $t \geq 0$, the sequence $(\mathbb{P}(W_t = k))_{k \in \mathbb{N}}$ is non-increasing on \mathbb{N} . We have:

$$\begin{aligned} \mathbb{E}[f(Y_t^{x+1})] &= \mathbb{E}[f(Y_t^x)] + e^{-qt} \sum_{k=0}^{\infty} \mathbb{P}(W_t = k) \mathbb{E}[f(Y_t^x + k + 1) - f(Y_t^x + k)] \\ &= (1 - e^{-qt} \mathbb{P}(W_t = 0)) \mathbb{E}[f(Y_t^x)] + e^{-qt} \sum_{k=1}^{\infty} (\mathbb{P}(W_t = k - 1) - \mathbb{P}(W_t = k)) \mathbb{E}[f(Y_t^x + k)] \\ &= \sum_{k=0}^{\infty} w(k) \mathbb{E}[f(Y_t^x + k)], \end{aligned}$$

where the sequence $(w(k))_{k \in \mathbb{N}}$ is defined in the statement of the lemma. It is easy to check that $\sum_{k=0}^{\infty} w(k) = 1$, and that the sequence $(w(k))_{k \in \mathbb{N}}$ is non-negative thanks to the observation above. For all $t \geq 0$, we define a random variable S_t such that S_t is independent of $(Y_t^x)_{t \geq 0}$ and that for all non-negative integer $\mathbb{P}(S_t = k) = w(k)$. This yields the equality in law $Y_t^{x+1} = Y_t^x + S_t$. The lemma follows by induction. \square

Let us turn to the study of Stein's factors associated to the negative binomial approximation. We begin by a lemma on the instantaneous probabilities of a Galton-Watson process with immigration.

Lemma 4.3 (Upper bound of the instantaneous probabilities of a GWI process). *Set $(X_t^x)_{t \geq 0}$ be a BDP $(p(r+k), k)_{k \in \mathbb{N}}$. We have:*

$$\begin{aligned} \sup_{x \in \mathbb{N}} \mathbb{P}(X_t^x = x) &\leq e^{-(1-p)t} \left(1 \wedge \frac{1}{p(1 - e^{-t})} \right), \quad t \geq 0, \quad p \in (0, 1), \quad r > 0, \\ \sup_{x \in \mathbb{N}} \mathbb{P}(X_t^x = x) &\leq \frac{1}{\sqrt{2e}} \left(\frac{1-p}{p(1 - e^{-(1-p)t})} \right)^{1/2} \frac{K(r)}{\sqrt{r}}, \quad t \geq 0, \quad p \in (0, 1), \quad r > \frac{1}{2}, \end{aligned}$$

with $K(r) = \sqrt{r} \Gamma(r - 1/2) / \Gamma(r)$.

The first upper bound is proved now, as a consequence of Lemma 7.1 and Lemma 4.2. The second upper bound is already known (Barbour et al. [2015]) and is based on results of Kendall [1948] and Phillips [1996].

Proof of Lemma 4.3. By Lemma 4.2,

$$\sup_{x \in \mathbb{N}} \mathbb{P}(X_t^x = x) \leq \sup_{x \in \mathbb{N}} \mathbb{P}(Z_{1,t} = x) \wedge \sup_{x \in \mathbb{N}} \mathbb{P}(X_t^0 = x), \quad t \geq 0,$$

where $Z_{1,t}$ is defined in Lemma 4.2. According to this lemma and to Lemma 7.1,

$$\begin{aligned}\mathbb{P}(Z_{1,t} = x) &\leq e^{-(1-p)t} (\mathbb{P}(W_t = k-1) - \mathbb{P}(W_t = k)) \\ &\leq e^{-(1-p)t} \left(1 \wedge \frac{C}{p(1-e^{-t})} \right),\end{aligned}\quad x \in \mathbb{N}, \quad t \geq 0.$$

On the other hand, by a result of Kendall [1948], cited as Lemma 2.2 in Barbour et al. [2015], it is known that for all $t \geq 0$, X_t^0 is distributed as a negative binomial distribution of parameters $(r, \theta_t(p))$, with

$$\theta_t(p) = 1 - \frac{1-p}{1-pe^{-(1-p)t}}.$$

Now Phillips [1996] shows that when X is distributed as a negative binomial distribution with parameters (r, θ) , and if $r > \frac{1}{2}$, then

$$\sup_{k \in \mathbb{N}} \mathbb{P}(X = k) \leq \frac{1}{\sqrt{2e}} \sqrt{\frac{1-\theta}{\theta}} \frac{K(r)}{\sqrt{r}}.$$

□

For the Stein factor associated with Lipschitz function, Theorem 3.6 and equation (16) yield

$$\sup_{f \in \text{Lip}(d)} \|g_f\|_\infty = \frac{1}{\sigma(1)} = \frac{1}{1-p},$$

recovering [Barbour et al., 2015, Theorem 1.1, equation (1.3)].

The following proposition on the second Stein factor associated to Lipschitz function improves on the known upper bounds.

Proposition 4.4 (Estimation of the second Stein's factor for Lipschitz function and NB-approximation). *Let $r > 0$ and $0 < p < 1$. For a real-valued function f on \mathbb{N} , let g_f be the (centered) solution to Stein's equation*

$$p(r+x) \partial g_f(x) + x \partial^* g_f(x) = f(x) - \int f d\text{NB}(r, p), \quad x \in \mathbb{N}.$$

Then,

$$\sup_{f \in \text{Lip}(d)} \|\partial g_f\|_\infty \leq \min \left\{ \frac{2}{3(1-p)}, \frac{D}{\sqrt{(r+2)p(1-p)}} \right\}, \quad D = 2 \frac{\sqrt{\pi}}{3\sqrt{e}} \simeq 0.72.$$

Proof. By application of Theorem 6.8 and formula (16), we find that

$$\sup_{f \in \text{Lip}(d)} \|\partial g_f\|_\infty = 2 \int_0^\infty e^{-2(1-p)t} \sup_{i \in \mathbb{N}} \mathbb{P}(X_{1,1,t}^i = i) dt,$$

where $(X_{1,1,t}^i)_{t \geq 0}$ is a BDP $(p(r+2+x), x)_{x \in \mathbb{N}}$. Applying Lemma 4.3,

$$\sup_{f \in \text{Lip}(d)} \|\partial g_f\|_\infty \leq 2 \int_0^\infty e^{-3(1-p)t} dt \wedge 2 \sqrt{\frac{1p}{p}} \frac{K(r+2)}{\sqrt{r+2}} \int_0^\infty \sqrt{e^{-2(1-p)t}} \sqrt{1-e^{-(1-p)t}} dt.$$

The function K is decreasing on $(\frac{1}{2}, \infty)$ and $\int_0^\infty \frac{e^{-2(1-p)t}}{\sqrt{1-e^{-(1-p)t}}} dt = \frac{4}{3(1-p)}$. Hence

$$\sup_{f \in \text{Lip}(d)} \|\partial g_f\|_\infty \leq \min \left\{ \frac{2}{3(1-p)}, \frac{D}{\sqrt{(r+2)p(1-p)}} \right\},$$

with $D = \frac{8K(2)}{3\sqrt{2e}} = \frac{4\Gamma(3/2)}{3\sqrt{e}} = \frac{2\sqrt{\pi}}{3\sqrt{e}}$.

□

The Proposition 4.4 might be compared to [Barbour et al., 2015, Theorem 1.1, equation (1.4)], which states the inequality

$$\sup_{f \in \text{Lip}(d)} \|\partial g_f\|_\infty \leq \min \left\{ \frac{2}{1-p}, \frac{1+p}{(1-p)^2}, \frac{1.5}{\sqrt{rp(1-p)^3}} \right\}. \quad (17)$$

We observe that:

- The numerical constant in front of $1/(1-p)$ is improved.
- As $\frac{D}{\sqrt{(r+2)p(1-p)}} \leq \frac{0.8}{\sqrt{rp(1-p)}}$ and $0.8 \leq \frac{1.5}{1-p}$, we have:

$$\frac{D}{\sqrt{(r+2)p(1-p)}} \leq \frac{1.5}{\sqrt{rp(1-p)^3}}.$$

Note that the proofs are similar up to the formula

$$\sup_{f \in \text{Lip}(d)} |\partial g_f(i)| = - \int_0^\infty \partial \partial^* P_t \mathbf{1}_i dt.$$

We then apply the second order intertwining formula, whereas Barbour et al. [2015] use another technique. In both cases, a bound of the type $\sup_i \mathbb{P}(Y_t^i = i)$ is needed, but not for the same process $(Y_t)_{t \geq 0}$.

Let us make two short remarks on the proof of Proposition 4.4: firstly, we do not use the upper bound

$$\sup_{x \in \mathbb{N}} \mathbb{P}(X_t^x = x) \leq e^{-(1-p)t} \frac{1}{p(1-e^{-t})},$$

that would make appear the hypergeometric function ${}_2F_1$. Secondly, the function K is bounded from below by a positive constant on $(\frac{1}{2}, \infty)$, hence by writing $K(r+2) \leq K(2)$ we do not lose the rate in r .

For the Stein factor associated to bounded functions, at the order 1 we find the following result.

Proposition 4.5 (Estimation of the first Stein factor for bounded functions and NB-approximation). *With the same assumptions as in Theorem 4.4, we have:*

$$\sup_{0 \leq f \leq 1} \|g_f\|_\infty \leq \frac{1}{2(1-p)} \wedge \frac{\sqrt{\pi}}{\sqrt{(r+1)p(1-p)}}.$$

We do not detail the proof which is very similar to the one of Proposition 4.4.

This result improves on a result of [Brown and Phillips, 1999, Lemma 3] which states

$$\sup_{0 \leq f \leq 1} \|g_f\|_\infty \leq \frac{1}{p \vee (1-p) \mathbf{1}_{r \geq 1}}.$$

We do not develop the case of the second Stein factor of bounded functions, where the upper bound given by Theorem 3.3 recovers the simple inequality

$$\sup_{0 \leq f \leq 1} \|\partial g_f\|_\infty \leq \sup_{f \in \text{Lip}(d_1)} \|\partial g_f\|_\infty.$$

Results about this factor can be found in [Brown and Xia, 2001, Theorem 2.10], in [Eichelsbacher and Reinert, 2008, example 2.12] for the case $r = 1$, and in [Brown and Phillips, 1999, Lemma 5].

If $\varphi : x \mapsto r + x$, $r \in \mathbb{N}$ and $\lambda \in (0, 1)$ then $\mathcal{I}_\varphi(\lambda) = \text{NB}(r, \lambda)$. The variable $W + r$ then represents the number of trials that are necessary to obtain r successes in a Bernoulli experiment with a random probability of gain.

To conclude this section, we observe that the Stein operator associated to a probability measure is not unique, and that resulting Stein's factors depend on the choice of the operator. When $r = 1$, we recover the geometric law as the invariant distribution, similarly to the forthcoming example. This is the choice of Eichelsbacher and Reinert [2008] to study the geometric distribution. In the next section we choose another Stein's operator.

4.3 The M/M/1 process and the geometric approximation

Let $(X_t^x)_{t \geq 0}$ be a BDP(α, β) with rates $\alpha(x) = \alpha$, $\beta(x) = \beta \mathbf{1}_{x \in \mathbb{N}^*}$ on \mathbb{N} . We suppose that $\rho := \frac{\alpha}{\beta} < 1$. We denote by $(P_t)_{t \geq 0}$ the associated semigroup. Its invariant distribution is the geometric law $\mathcal{G}(\rho)$ with pointwise probabilities $p(k) = (1 - \rho)\rho^k$ for $k \in \mathbb{N}$. Notice that this is the definition of the geometric law with support \mathbb{N} and not \mathbb{N}^* . Let us choose $u(x) = r^x$, $v(x) = q^x$ for $x \in \mathbb{N}$ with $r > 0$, $q \geq 1$. Theorem 2.1 gives rise to a Feynman-Kac semigroup composed of a birth-death semigroup $(P_{u,t})_{t \geq 0}$ with rates (α_u, β_u) and a potential V_u , which are defined as

$$\alpha_u(x) = r\alpha, \quad \beta_u(x) = \frac{1}{r}\beta, \quad V_u(x) = (1 - r)\alpha + \left(1 - \frac{1}{r}\mathbf{1}_{x \in \mathbb{N}^*}\right)\beta, \quad x \in \mathbb{N}.$$

The semigroup $(P_{u,t}^{V_u})_{t \geq 0}$ is still a semigroup associated to a M/M/1 queue, only with modified rates. The potential V_u , while non-constant, is non-increasing on \mathbb{N} . By Theorem 2.2, we find a Feynman-Kac semigroup $(P_{u,*v,t}^{V_u,*v})_{t \geq 0}$ where $(P_{u,*v,t})_{t \geq 0}$ is again a semigroup corresponding to a M/M/1 queue. The rates and potential are defined on \mathbb{N} as

$$\begin{aligned} \alpha_{u,*v}(x) &= qr\alpha, \quad \beta_{u,*v}(x) = \frac{1}{qr}\beta \mathbf{1}_{x \in \mathbb{N}^*}, \quad x \in \mathbb{N}, \\ V_{u,*v}(x) &= (1 - qr)\alpha + \left(1 - \frac{1}{qr}\right)\beta, \quad x \in \mathbb{N}^*, \quad V_{u,*v}(0) = \alpha - (1 + q)r\alpha + \beta. \end{aligned}$$

Remark that, in contrast with the general case of Theorem 2.2, the semigroup $(P_{u,*v,t})_{t \geq 0}$ is again a birth-death semigroup. This is due to the fact that V_u is constant on \mathbb{N}^* . The potential V_u is not constant on \mathbb{N} , which prevents us to apply Theorem 2.5, but it is almost constant which explains heuristically why we find again a birth-death process when applying Theorem 2.2.

Set $\sigma(u, *v) = \inf_{x \in \mathbb{N}} V_{u,*v}(x) = \min(V_{u,*v}(0), V_{u,*v}(1))$. A few calculations show that

$$\max \{ \sigma(u, *v) \mid u(x) = r^x, v(x) = q^x, r > 0, q \geq 1 \} = (\sqrt{\beta} - \sqrt{\alpha})^2,$$

and the $\arg \max$ is realized for all $r \leq \sqrt{\beta/\alpha} = \rho^{-1}$ and $q = \rho^{-1}/r$. This means that there is a range of choice for the parameters (r, q) allowing to recover the spectral gap $(\sqrt{\beta} - \sqrt{\alpha})^2$ of the process in the convergence of Theorem 2.8. However, contrary to the two preceding examples, notice that the second order intertwining does not allow to improve on the spectral gap and that the rate of convergence in the distance $\zeta_{u,*v}$ is the same as the rate of convergence in the Wasserstein distance W_{d_u} for the best choices of u, v .

This example is maybe the most important because, in contrast with the two previous processes, the M/M/1 queue is not known to satisfy a Mehler formula of the type (15), which would make it rather difficult to differentiate directly. A Mehler-like formula can nevertheless be deduced from Theorem 2.1: choosing $u = 1$ in this theorem, we derive

$$\mathbb{E}[f(X_t^{x+1}) - f(X_t^x)] = \mathbb{E} \left[e^{-\int_0^t V(X_s^x) ds} (f(X_t^x + 1) - f(X_t^x)) \right],$$

where $(X_t^x)_{t \geq 0}$ is M/M/1 process starting from x and $V(x) = \beta \mathbf{1}_{x=0}$. As a consequence, if B_t is a Bernoulli random variable verifying

$$\mathbb{P}(B_t = 1 \mid (X_s^x)_{s \leq t}) = e^{-\int_0^t V(X_s^x) ds}, \quad t \geq 0,$$

then,

$$\mathbb{E}[f(X_t^{x+1})] = \mathbb{E}[(f(X_t^x + B_t))], \quad t \geq 0,$$

and by induction there exists a random variable Y_t^x such that

$$\mathbb{E}[f(X_t^x)] = \mathbb{E}[(f(X_t^0 + Y_t^x))], \quad t \geq 0.$$

This formula seems to be new (even if the instantaneous distribution of the M/M/1 process is known, see Baccelli and Massey [1989]). Unfortunately, the random variable Y_t^x is not independent from X_t^0 and this makes this formula less powerful than (15). This approach is generalizable for every BDP with constant birth rate (so that the processes $(X_{1,t})_{t \geq 0}$ and $(X_t)_{t \geq 0}$ have the same law).

As in the preceding examples, we state a lemma related to the instantaneous probabilities of the modified process before turning to the Stein factors for geometric approximation.

Lemma 4.6 (Upper bound of the instantaneous probabilities of a M/M/1 queue). *Let $(Y_t)_{t \geq 0}$ be a M/M/1 queue with rates $(\lambda, \lambda \mathbf{1}_{\mathbb{N}^*})$. Then for all $t \geq 0$,*

$$\sup_{i \in \mathbb{N}^*} \mathbb{P}(Y_t^i = i) \leq \frac{1}{\sqrt{\lambda t}}.$$

Proof. Let us consider the BDP $(\tilde{Y}_t)_{t \geq 0}$ with rates $(1, \mathbf{1}_{x \in \mathbb{N}^*})$. Then for all $t \geq 0$, the equality in law $Y_t = \tilde{Y}_{\lambda t}$ holds, hence it is enough to prove that $\sup_{i \in \mathbb{N}^*} \mathbb{P}(\tilde{Y}_t^i = i) \leq \frac{1}{\sqrt{t}}$. By [Abate et al., 1991, Corollary 1 (d)], the sequence $(\mathbb{P}(\tilde{Y}_t^i = i))_{i \geq 0}$ is non-increasing for every $t \geq 0$. Hence $\sup_{i \in \mathbb{N}} \mathbb{P}(\tilde{Y}_t^i = i) = \mathbb{P}(\tilde{Y}_t^0 = 0)$. By [Abate et al., 1991, formula (9) and Corollary 2 (a)],

$$\mathbb{P}(\tilde{Y}_t^0 = 0) = \sum_{j=1}^{\infty} \frac{j}{t} \mathbb{P}(Z_t^0 = j) = \frac{1}{t} \mathbb{E}[Z_t^0 \mathbf{1}_{Z_t^0 > 0}],$$

where $(Z_t^0)_{t \geq 0}$ is a birth-death process with constant birth rate 1 and constant death rate 1 on the whole integer line \mathbb{Z} ; namely this is the continuous-time simple random-walk. This process can be represented as

$$\forall t \geq 0, \quad Z_t^0 = N_t^+ - N_t^-,$$

where $(N_t^+)_{t \geq 0}$ and $(N_t^-)_{t \geq 0}$ are two independent Poisson processes with intensity 1. So, using that N^1 and N^2 have the same law and Cauchy-Schwarz's inequality

$$\begin{aligned} \mathbb{E}[Z_t^0 \mathbf{1}_{Z_t^0 > 0}] &= \mathbb{E}[(N_t^+ - N_t^-) \mathbf{1}_{N_t^+ > N_t^-}] = \mathbb{E}[(N_t^- - N_t^+) \mathbf{1}_{N_t^- > N_t^+}] \\ &= \frac{1}{2} \mathbb{E}[|N_t^+ - N_t^-|] \leq \frac{1}{2} \text{Var}(N_t^+ - N_t^-)^{1/2} = \sqrt{\frac{t}{2}}. \end{aligned}$$

This yields

$$\sup_{i \in \mathbb{N}^*} \mathbb{P}(\tilde{Y}_t^i = i) \leq \mathbb{P}(\tilde{Y}_t^0 = 0) \leq \frac{1}{\sqrt{2t}},$$

which achieves the proof. \square

Up to the knowledge of the authors, Stein's factors associated to the Wasserstein distance have not been studied yet. The following proposition provides upper bounds on these factors.

Proposition 4.7 (Estimation of the Stein's factors for Lipschitz function and geometric approximation). *For all $0 < \alpha < \beta$, set $u(x) = q^x$ on \mathbb{N} with $q = \sqrt{\frac{\beta}{\alpha}} = \rho^{-1/2}$. Then,*

$$\begin{aligned} \sup_{f \in \text{Lip}(d_u)} \left\| \frac{gf}{u} \right\|_{\infty} &= \frac{1}{\sigma(u)} = \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})^2}, \\ \sup_{f \in \text{Lip}(d_u)} \|\partial_v g_f\|_{\infty} &\leq \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})^2} \left(1 + \sqrt{\frac{\alpha}{\beta}} \min \left\{ 1, \frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}} (\sqrt{\beta} - \sqrt{\alpha}) - 1 \right\} \right). \end{aligned}$$

Proof of Proposition 4.7. By application of Theorem 3.6, one has immediately the first equation. By Theorem 3.7 with $u(x) = q^x$, $q = \rho^{-1/2} = \sqrt{\frac{\beta}{\alpha}}$, we have:

$$\sup_{f \in \text{Lip}(d_u)} \|\partial_v g_f\|_{\infty} \leq \int_0^{\infty} e^{-(\sqrt{\beta} - \sqrt{\alpha})^2 t} \left(1 - \sqrt{\frac{\alpha}{\beta}} + 2\sqrt{\frac{\alpha}{\beta}} \sup_{i \in \mathbb{N}^*} \mathbb{P}(X_{1,*u,t}^i = i) \right) dt,$$

where $(X_{1,*u,t}^i)_{t \geq 0}$ is a M/M/1 queue with rates $(\sqrt{\alpha\beta}, \sqrt{\alpha\beta} \mathbf{1}_{\mathbb{N}^*})$. On the one hand, this yields directly

$$\sup_{f \in \text{Lip}(d_u)} \|\partial_v g_f\|_{\infty} \leq \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})^2} \left(1 + \sqrt{\frac{\alpha}{\beta}} \right).$$

On the other hand, as a consequence of Lemma 4.6, one has

$$\begin{aligned} \sup_{f \in \text{Lip}(d_u)} \|\partial_v g_f\|_{\infty} &\leq \int_0^{\infty} e^{-(\sqrt{\beta} - \sqrt{\alpha})^2 t} \left(1 - \sqrt{\frac{\alpha}{\beta}} + 2\frac{1}{(\alpha\beta)^{1/4}} \sqrt{\frac{\alpha}{\beta}} \frac{1}{\sqrt{t}} \right) dt \\ &= \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})^2} \left(1 - \sqrt{\frac{\alpha}{\beta}} \right) + \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})} \sqrt{\frac{\alpha}{\beta}} \frac{2}{(\alpha\beta)^{1/4}} \int_0^{\infty} e^{-t} \frac{dt}{\sqrt{t}} \\ &= \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})^2} \left(1 - \sqrt{\frac{\alpha}{\beta}} \right) + \frac{1}{(\sqrt{\beta} - \sqrt{\alpha})} \sqrt{\frac{\alpha}{\beta}} \frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}}. \end{aligned}$$

□

Remark 4.8 (On the best upper bound). *The expression $\frac{2\sqrt{\pi}}{(\alpha\beta)^{1/4}}(\sqrt{\beta} - \sqrt{\alpha}) - 1$ is smaller than 1 as soon as*

$$\frac{\sqrt{\beta} - \sqrt{\alpha}}{(\alpha\beta)^{1/4}} < \frac{1}{\sqrt{\pi}},$$

so there is a range of values of the parameters α and β , for example if they are close to each other, for which the factor inside the min is actually a better upper bound than 1.

We now turn to the subject of the mixture of geometric laws. Set $\varphi = 1$ and $\rho < 1$, then $\mathcal{I}_{\varphi}(\rho) = \mathcal{G}(\rho)$. We choose $u(k) = q^k$ on \mathbb{N} , hence $d_u(x, y) = |q^x - q^y|/|q - 1|$. The preceding theorem put together with Theorem 3.12 gives for $q = \rho^{-1/2}$ and in the case where $\rho' < \sqrt{\rho}$,

$$W_{d_u}(\mathcal{G}(\rho), \mathcal{G}(\rho')) \leq |\rho - \rho'| \times \frac{1}{(1 - \sqrt{\rho})^2} \times \frac{1 - \rho'}{\sqrt{\rho} - \rho'}.$$

The case $\rho' > \sqrt{\rho}$ is similar.

By the same reasoning as the one used in the proof of Theorem 3.14, for a random variable R such that $\mathbb{E}[R] = \rho$, and a random variable such that $\mathcal{L}(W|R) = \mathcal{G}(R)$, we have the inequality:

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{G}(\rho)) \leq \sup_{f \in \mathcal{F}} \|\partial_u g_f\|_{\infty} \mathbb{E}[(\rho - R)d_u(W + 1, G + 1)],$$

where $G \sim \mathcal{G}(\rho)$. Let $G' \sim \mathcal{G}(\rho')$. With the interpretation of the geometric laws as the number of repetitions of a binary experiment before the first success, it is easy to find a coupling such that a.s. $G \leq G'$ when $\rho \leq \rho'$. This yields

$$\mathbb{E}[d_u(G, G')] = \frac{1}{|1-q|} \left| \frac{1-\rho}{1-q\rho} - \frac{1-\rho'}{1-q\rho'} \right| = \frac{|\rho - \rho'|}{|(1-q\rho)(1-q\rho')|}.$$

Hence, if a.s. $R < \frac{1}{q}$, by Remark 3.15:

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{G}(\rho)) \leq \sup_{f \in \mathcal{F}} \|\partial_u g_f\|_{\infty} \frac{q}{1-q\rho} \mathbb{E} \left[\frac{|\rho - R|^2}{(1-qR)} \right].$$

Finally, by taking $q = \rho^{-1/2}$, one finds that for two random variables R, S such that $\mathbb{E}[R] = \rho$ and a.s. $R < \frac{1}{\sqrt{\rho}}$, and $\mathcal{L}(W|R) = \mathcal{G}(R)$, the following upper bound holds:

$$d_{\mathcal{F}}(\mathcal{L}(W), \mathcal{G}(\rho)) \leq \frac{1 + \frac{1}{\sqrt{\rho}}}{(1 - \sqrt{\rho})^3} \mathbb{E} \left[\frac{|\rho - R|^2}{(1 - \frac{R}{\sqrt{\rho}})} \right].$$

4.4 Another example

Let us consider the BDP(α, β) with $\alpha(x) = x + 2, \beta(x) = x^2$ on \mathbb{N} . Its invariant measure is a Poisson size-biased type distribution, defined as

$$\pi(x) = \frac{1}{2e} \frac{(x+1)}{x!}, \quad x \in \mathbb{N}.$$

Here size-biased means that if $X \sim \pi$ and $Y \sim \mathcal{P}(1)$ then:

$$\mathbb{P}(X = x) = \frac{\mathbb{E}[(Y+1)\mathbf{1}_{Y=x}]}{\mathbb{E}[(Y+1)]} = \frac{(x+1)\mathbb{P}(Y=x)}{\sum_{j \geq 0} (j+1)\mathbb{P}(Y=j)}, \quad x \in \mathbb{N}.$$

Choosing the weight u such that $u(x+1)/u(x) = (x+1)/(x+3)$ for all $x \in \mathbb{N}$, i.e. for example

$$u(x) = \frac{1}{(x+1)(x+2)}, \quad x \in \mathbb{N},$$

we find that V_u is constant. By Theorem 2.5 with $v = 1$, we have an intertwining with potential $V_{u,v}(x) = 2x+1$ on \mathbb{N} . Moreover, by Theorem 2.8, we have convergence of the semigroup towards π in the distance $\zeta_{u,1}$ at rate 1.

The three next sections are devoted to the omitted proofs of the previous results.

5 Proofs of Section 2

5.1 First order intertwining for the backward gradient ∂_u^*

First of all, let us state the analogous of Theorem 2.1 for the backward gradient ∂^* . Let $(P_{*u,t})_{t \geq 0}$ be the birth-death semigroup associated to the generator L_{*u} , where for all non-negative or bounded function $f : \mathbb{N} \rightarrow \mathbb{R}$ and $x \in \mathbb{N}$,

$$\begin{aligned} L_{*u}f &= \alpha_{*u} \partial f + \beta_{*u} \partial^* f, & V_{*u} &= \overleftarrow{\alpha} - \alpha_{*u} + \beta - \beta_{*u}, \\ \alpha_{*u}(x) &= \frac{u(x+1)}{u(x)} \alpha(x), & \beta_{*u}(x) &= \frac{u(x-1)}{u(x)} \beta(x-1) \mathbf{1}_{x \in \mathbb{N}^*}. \end{aligned}$$

The potential V_{*u} can be rewritten under the compacted form $V_{*u} = \partial_u^* \left(\overrightarrow{u} \alpha - u \beta \right)$. We can also notice that $V_{*u} = \overleftarrow{V_u}$ on \mathbb{N}^* .

Theorem 5.1 (First-order intertwining relation for the backward gradient). *If V_{*u} is bounded from below, then for every real-valued function on \mathbb{N} such that $\|\partial_u^* f\|_\infty < +\infty$, and for all $t \geq 0$,*

$$\partial_u^* P_t f = P_{*u,t}^{V_{*u}} \partial_u^* f. \quad (18)$$

Let us call $(X_{*u,t}^x)_{t \geq 0}$ the birth-death process of generator L_{*u} such that $X_{*u,0}^x = x$. The process $(X_{*u,t}^x)_{t \geq 0}$ is not irreducible, although it is indecomposable, i.e. it possesses only one recurrent class. Indeed if $x \in \mathbb{N}^*$ then $(X_{*u,t}^x)_{t \geq 0}$ never visits the state 0 as $\beta_{*u}(1) = 0$ and if $x = 0$ the process $(X_{*u,t}^0)_{t \geq 0}$ leaves 0 almost surely.

Proof of Theorem 5.1. The core of the proof relies on the intertwining relation at the level of generators:

$$\partial_u^* L f = L_{*u} \partial_u^* f - V_{*u} \partial_u^* f, \quad (19)$$

which derives by easy computations. The intertwining at the level of the semigroups follows by the same arguments as in the proof of Theorem 2.1 of Chafai and Joulin [2013]. We briefly recall these arguments. For all $s \in [0, t]$ let us set $J(s) = P_{*u,s}^{V_{*u}}(\partial_u^* P_{t-s} f)$. If the function $\partial_u^* P_{t-s} f$ is bounded on \mathbb{N} , then the Kolmogorov equations (6) for the Feynman-Kac semigroup $(P_{*u,t}^{V_{*u}})_{t \geq 0}$ hold and

$$J'(s) = P_{*u,s}^{V_{*u}}((L_{*u} - V_{*u})\partial_u^* P_{t-s} f - \partial_u^* L P_{t-s} f).$$

Thanks to the formula (19) this gives $J'(s) = 0$. Hence $J(0) = J(t)$ which is exactly the identity (18).

Let us show that $\partial_u^* P_{t-s} f$ is bounded on \mathbb{N} . Indeed, recall that $V_{*u}(x+1) = V_u(x)$ on \mathbb{N} . Furthermore $\partial_u^* f(x+1) = \partial_u f(x)$ on \mathbb{N} . Hence V_u and $\partial_u f$ are bounded on \mathbb{N} , which implies that $\partial_u P_{t-s} f$ is bounded (Chen [1996]). For all positive integer $\partial_u P_{t-s} f(x) = \partial_u^* P_{t-s} f(x+1)$, so $\partial_u^* P_{t-s} f$ is bounded. \square

5.2 Alternative proof of first order intertwining theorems

This section aims to give a sample path interpretation of the first order intertwining relations (7) and (18), at least in a particular case. It is independent of the other sections.

We focus on the case where the weight is $u = 1$ with non-increasing birth rates $(\alpha(x))_{x \in \mathbb{N}}$ and non-decreasing death rates $(\beta(x))_{x \in \mathbb{N}}$. When intertwining the birth-death semigroup with the forward gradient ∂ , one obtains a new birth-death semigroup with shifted birth rate and unchanged death rate

$$\alpha_1 = \vec{\alpha}, \quad \beta_1 = \beta,$$

whereas when intertwining the birth-death semigroup with the backward gradient ∂^* , one obtains a new birth-death semigroup with shifted death rate and unchanged birth rate:

$$\alpha_{*1} = \alpha, \quad \beta_{*1} = \overleftarrow{\beta}.$$

In order to explain this fact, we will give a probabilistic proof of the formulae (7) and (18). Recall that for all real-valued bounded functions on \mathbb{N} and $x \in \mathbb{N}$,

$$\begin{aligned} \partial P_t f(x) &= \mathbb{E}[f(X_t^{x+1}) - f(X_t^x)], \\ \partial^* P_t f(x+1) &= \mathbb{E}[f(X_t^x) - f(X_t^{x+1})]. \end{aligned}$$

At time $t = 0$, $X_t^{x+1} = X_t^x + 1$. We construct a process $(S_t)_{t \geq 0}$ such that for all $t \geq 0$, $X_t^{x+1} = X_t^x + S_t$ and $S_t \in \{0, 1\}$. If for a time t , $S_t = 0$, then we choose the sticking coupling between $(X_{t+s}^x)_{s \geq 0}$ and $(X_{t+s}^{x+1})_{s \geq 0}$ (i.e. the process $(S_t)_{t \geq 0}$ is absorbed in 0). If $S_t = 1$, it is natural to construct the following coupling:

1. with rate $\alpha(X_t^x + 1) = \alpha(X_t^{x+1})$, X_t^x and X_t^{x+1} jump upwards together and S_t remains equal to 1,

2. with rate $\beta(X_t^x) = \beta(X_t^{x+1} - 1)$, X_t^x and X_t^{x+1} jump downwards together and S_t remains equal to 1,
3. with rate $\alpha(X_t^x) - \alpha(X_t^x + 1) = \alpha(X_t^{x+1} - 1) - \alpha(X_t^{x+1})$, X_t^x jumps upwards, X_t^{x+1} does not jump and S_t jumps from 1 to 0,
4. with rate $\beta(X_t^{x+1}) - \beta(X_t^x) = \beta(X_t^{x+1}) - \beta(X_t^{x+1} - 1)$, X_t^{x+1} jumps downwards, X_t^x does not jump and S_t jumps from 1 to 0.

This implies in particular that for all $t \geq 0$ the process S_t jumps from 1 to 0 with rate

$$\alpha(X_t^x) - \alpha(X_t^x + 1) + \beta(X_t^{x+1}) - \beta(X_t^x) = V_1(X_t^x) = V_{*1}(X_t^{x+1}).$$

Moreover, conditionally to $\{S_t = 1\}$, $(X_t^x)_{t \geq 0}$ evolves as a BDP($\vec{\alpha}, \beta$) and $(X_t^{x+1})_{t \geq 0}$ evolves as a BDP($\alpha, \vec{\beta}$). Indeed, as long as $S_t = 1$, the steps (3) and (4) do not occur.

To exploit rigorously the preceding facts, let us introduce the BDP($\vec{\alpha}, \beta$) starting from x denoted by $(X_{1,t}^x)_{t \geq 0}$, whose standard filtration is $(\mathcal{F}_t)_{t \geq 0}$. The processes $(X_t^x)_{t \geq 0}$ and $(X_{1,t}^x)_{t \geq 0}$, as well as $(X_t^{x+1})_{t \geq 0}$ and $(X_{1,t}^x + 1)_{t \geq 0}$, can be coupled as follows :

1. Let E be an exponential with parameter 1 and T such that $T = \inf\{t \geq 0, \int_0^t V(X_{1,s}^x) ds > E\}$.
2. Set $S_t = 1$ if $t < T$ and $S_t = 0$ otherwise.
3. Set $X_t^x = X_{1,t}^x$ for $t \leq T$.
4. At time T , sample a random variable Z satisfying to

$$\begin{aligned} \mathbb{P}(Z = X_{1,T}^x + 1 \mid \mathcal{F}_T) &= \frac{\alpha(X_{1,T}^x) - \alpha(X_{1,T}^x + 1)}{V(X_{1,T}^x)}, \\ \mathbb{P}(Z = X_{1,T}^x \mid \mathcal{F}_T) &= \frac{\beta(X_{1,T}^x + 1) - \beta(X_{1,T}^x)}{V_1(X_{1,T}^x)}. \end{aligned}$$

5. Let evolve the process $(X_t^x)_{t \geq T}$ as a BDP(α, β) starting from Z .

The coupling $(X_t^x, X_{1,t}^x, S_t)_{t \geq 0}$ satisfy to

$$X_t^x \mathbf{1}_{S_t=1} = X_{1,t}^x \mathbf{1}_{S_t=1}, \quad X_t^{x+1} \mathbf{1}_{S_t=1} = (X_{1,t}^x + 1) \mathbf{1}_{S_t=1}, \quad \mathbb{P}(S_t = 1 \mid (X_{1,s}^x)_{0 \leq s \leq t}) = e^{-\int_0^t V_1(X_{1,s}^x) ds}.$$

This allows to find back the formula (7):

$$\begin{aligned} \partial P_t f(x) &= \mathbb{E}[f(X_t^{x+1}) - f(X_t^x)] = \mathbb{E}[(f(X_t^{x+1}) - f(X_t^x)) \mathbf{1}_{S_t=1}] \\ &= \mathbb{E}\left[(f(X_{1,t}^x + 1) - f(X_{1,t}^x)) e^{-\int_0^t V_1(X_{1,s}^x) ds}\right] = P_{1,t}^{V_1}(\partial f)(x). \end{aligned}$$

Similarly it is possible to construct a coupling $(X_t^{x+1}, X_{*1,t}^{x+1}, S_t)_{t \geq 0}$ such that $(X_{*1,t}^{x+1})_{t \geq 0}$ is a BDP($\alpha, \vec{\beta}$) starting from $x + 1$ and satisfying to

$$X_t^{x+1} \mathbf{1}_{S_t=1} = X_{*1,t}^{x+1} \mathbf{1}_{S_t=1}, \quad X_t^x \mathbf{1}_{S_t=1} = (X_{*1,t}^{x+1} - 1) \mathbf{1}_{S_t=1}, \quad \mathbb{P}(S_t = 1 \mid (X_{*1,s}^{x+1})_{0 \leq s \leq t}) = e^{-\int_0^t V_{*1}(X_{*1,s}^{x+1}) ds},$$

leading to the formula (18):

$$\begin{aligned} \partial^* P_t f(x + 1) &= \mathbb{E}[f(X_t^x) - f(X_t^{x+1})] = \mathbb{E}[(f(X_t^x) - f(X_t^{x+1})) \mathbf{1}_{S_t=1}] \\ &= \mathbb{E}\left[(f(X_{*1,t}^{x+1} - 1) - f(X_{*1,t}^{x+1})) e^{-\int_0^t V_{*1}(X_{*1,s}^{x+1}) ds}\right] = P_{*1,t}^{V_{*1}}(\partial^* f)(x + 1). \end{aligned}$$

It is interesting to remark that conversely, the intertwining formula (7) can in certain cases yield a coupling between $(X_t^x)_{t \geq 0}$ and $(X_t^{x+1})_{t \geq 0}$. The proof of Lemma 4.2 above is based on this idea.

5.3 Proof of Theorems 2.2 and 2.5

Proof of Theorem 2.2. Let us begin by showing the following intertwining relation at the level of the generators:

$$\partial_v^* \partial_u L f = L_{u,*v} \partial_v^* \partial_u f - V_{u,*v} \partial_v^* \partial_u f.$$

By application of Theorem 2.1 and Theorem 5.1 we find that

$$\begin{aligned} \partial_v^* (\partial_u L f) &= \partial_v^* (L_u (\partial_u f) - V_u \partial_u f) \\ &= (L_u)_{*v} \partial_v^* \partial_u f - (V_u)_{*v} \partial_v^* \partial_u f + \partial_v^* (-V_u \partial_u f), \end{aligned}$$

where $(L_u)_{*v}$ and $(V_u)_{*v}$ stand for the generator, respectively the potential, obtained by intertwining the $\text{BDP}(\alpha_u, \beta_u)$ and the ∂_{*v} gradient. The generator $(L_u)_{*v}$ is the generator of a $\text{BDP}((\alpha_u)_{*v}, (\beta_u)_{*v})$ such that for all $x \in \mathbb{N}$,

$$\begin{aligned} (\alpha_u)_{*v}(x) &= \frac{v(x+1)}{v(x)} \alpha_u(x) = \frac{v(x+1)}{v(x)} \frac{u(x+1)}{u(x)} \alpha(x+1) \\ (\beta_u)_{*v}(x) &= \frac{v(x-1)}{v(x)} \beta_u(x-1) = \frac{v(x-1)}{v(x)} \frac{u(x-2)}{u(x-1)} \beta(x-1) \mathbf{1}_{x \in \mathbb{N}^*}. \end{aligned}$$

The potential $(V_u)_{*v}$ writes on \mathbb{N}

$$(V_u)_{*v}(x) = \alpha_u(x-1) \mathbf{1}_{x \in \mathbb{N}^*} - (\alpha_u)_{*v}(x) + \beta_u(x) - (\beta_u)_{*v}(x).$$

The next step is to rewrite the expression $\partial_v^* (-V_u \partial_u f)$ in terms of $\partial_v^* \partial_u f$. Let us denote $g = \partial_u f$ in the following lines. For every $x \in \mathbb{N}^*$, $\partial_u^* (fg)(x) = f(x) \partial_u^* g(x) + \partial_u^* f(x) g(x-1)$ and $f(x) = -\sum_{k=0}^x u(k) \partial_u^* f(k)$ so that

$$\begin{aligned} \partial_v^* (-V_u g)(x) &= -V_u(x) \partial_v^* g(x) - \partial_v^* V_u(x) g(x-1) \\ &= -V_u(x) \partial_v^* g(x) + \partial_v^* V_u(x) \sum_{k=0}^{x-1} v(k) \partial_v^* g(k) \\ &= \partial_v^* V_u(x) \sum_{k=0}^{x-1} v(k) (\partial_v^* g(k) - \partial_v^* g(x)) - \left(V_u(x) - \left(\sum_{k=0}^{x-1} v(k) \right) \partial_v^* V_u(x) \right) \partial_v^* g(x). \end{aligned}$$

Besides, $\partial_v^* (-V_u g)(0) = \frac{1}{v_0} V_u(0) g(0) = -V_u(0) \partial_v^* g(0)$. We do indeed find $\partial_v^* \partial_u L = (L_{u,*v} - V_{u,*v}) \partial_v^* \partial_u$ with

$$\begin{aligned} L_{u,*v} f(x) &= (L_u)_{*v} f(x) + \partial_v^* V_u(x) v(x-1) (f(x-1) - f(x)) \\ &\quad + \partial_v^* V_u(x) \left(\sum_{j=0}^{x-2} v(j) \right) \sum_{k=0}^{x-2} \frac{v(k)}{\left(\sum_{j=0}^{x-2} v(j) \right)} (f(k) - f(x)) \\ V_{u,*v}(x) &= (V_u)_{*v}(x) + V_u(x) - \left(\sum_{k=0}^{x-1} v(k) \right) \partial_v^* V_u(x). \end{aligned}$$

The generator $L_{u,*v}$ has a birth-death component and a component making the process at point x jumping on the set $\{0, \dots, x-2\}$. The birth rates are $\alpha_{u,*v} = (\alpha_u)_{*v}$. The death rates come from $(L_u)_{*v}$ and from the term $\partial_v^* V_u(x) v(x-1) (f(x-1) - f(x))$, so that

$$\beta_{u,*v}(x) = (\beta_u)_{*v} + \partial_v^* V_u(x) v(x-1) \mathbf{1}_{x \in \mathbb{N}^*}.$$

Remembering that $V_u(x) = \alpha(x) - \alpha_u(x) + \beta(x+1) - \beta_u(x)$ we get that for all positive integer x

$$\begin{aligned} (V_u)_{*v}(x) + V_u(x) &= \alpha(x) + \alpha_u(x-1) - (\alpha_u(x) + (\alpha_u)_{*v}(x)) + \beta(x+1) - (\beta_u)_{*v}(x) \\ &= \left(1 + \frac{u(x)}{u(x-1)} \right) \alpha(x) - \left(1 + \frac{v(x+1)}{v(x)} \right) \frac{u(x+1)}{u(x)} \alpha(x+1) \\ &\quad + \beta(x+1) - \frac{v(x-1)}{v(x)} \frac{u(x-2)}{u(x-1)} \beta(x-1), \end{aligned}$$

and

$$\begin{aligned}
(V_u)_{*v}(0) + V_u(0) &= -(\alpha_u)_{*v}(0) + V_u(0) = \alpha(0) - (\alpha_u(0) + (\alpha_u)_{*v}(0)) + \beta(1) \\
&= \alpha(0) - \left(1 + \frac{v(1)}{v(0)}\right) \frac{u(1)}{u(0)} \alpha(1) + \beta(1).
\end{aligned}$$

The same reasoning as in the proof of Theorem 5.1 allows to deduce the relation at the level of the semigroups from the relation at the level of the generators, provided that we can show that for all $t \geq 0$ the function $\partial_v^* \partial_u P_t f$ is bounded on \mathbb{N} . It is the case; indeed, by Theorem 2.1, $\partial_u P_t f = P_{u,t}^{V_u} \partial_u f$ is bounded and $\partial_v^* |\partial_u P_t f| \leq \frac{2}{\inf_{x \in \mathbb{N}} v(x)} |P_{u,t}^{V_u} \partial_u f|$. \square

Proof of Theorem 2.5. Surprisingly, Theorem 2.5 cannot be deduced from Theorem 2.2 when $u \neq 1$. However, its proof goes along the same lines as the proof of Theorem 2.2, only easier because $\partial_v \partial_u (V_u \partial_u f) = V_u \partial_v \partial_u f$, so that the intertwining relation at the level of the generators follows directly. \square

6 Proofs of Section 3

The semigroup representation (13) of the solution of Stein's equation g_f can be rewritten as:

$$\vec{g}_f = -u \int_0^\infty \partial_u P_t f dt, \quad (20)$$

$$\partial g_f = u \int_0^\infty \partial_u \partial^* P_t f dt, \quad (21)$$

$$\overrightarrow{\partial g}_f = -u \int_0^\infty \partial_u \partial P_t f dt. \quad (22)$$

The left-hand side of an intertwining relation between a weighted gradient and a birth-death semigroup appears under the integral. This fact suggests to apply the intertwining relations shown previously. However, it leads to sharper results to first identify the function $f \in \mathcal{F}$ that realizes the maximum in the pointwise Stein's factors

$$\max_{f \in \mathcal{F}} |g_f(i)|, \quad \max_{f \in \mathcal{F}} |\partial g_f(i)|,$$

for every $i \in \mathbb{N}$. This first step is based on Lemma 6.1 and Lemma 6.2 below. Indeed, Lemma 6.1 gives an alternative formulation of the solution of Stein's equation.

Lemma 6.1 ([Brown and Xia, 2001, Lemma 2.3]). *For all $i \in \mathbb{N}$, let us define $g_j := g_{1_j}$ and*

$$e_i^+ = \frac{1}{\alpha(i)\pi(i)} \sum_{k=0}^i \pi(k), \quad i \in \mathbb{N} \quad \quad e_i^- = \frac{1}{\beta(i)\pi(i)} \sum_{k=i}^\infty \pi(k), \quad i \in \mathbb{N}^*.$$

Then, for all $i \in \mathbb{N}^, j \in \mathbb{N}$,*

$$g_j(i) = \pi(j)(-e_{i-1}^+ \mathbf{1}_{i \leq j} + e_i^- \mathbf{1}_{i \geq j+1}) \quad (23)$$

$$\partial g_j(i) = \pi(j) \left((e_{i-1}^+ - e_i^+) \mathbf{1}_{j \geq i+1} + (e_{i+1}^- + e_{i-1}^+) \mathbf{1}_{i=j} + (e_{i+1}^- - e_i^-) \mathbf{1}_{j \leq i-1} \right). \quad (24)$$

Lemma 6.2 ([Brown and Xia, 2001, Lemma 2.4]). *If $V_1 \geq 0$ then (e_i^+) is non-decreasing and (e_i^-) is non-increasing.*

6.1 Approximation in total variation distance

We begin by describing the argmax of the pointwise quantities. To the knowledge of the authors, equation (25) is not explicitly stated in preceding works. Equation (26) is proved in Brown and Xia [2001]. We briefly recall the arguments used for the sake of completeness.

Lemma 6.3 (Argmax of the pointwise Stein's factor). *For all $i \in \mathbb{N}$,*

$$g_{\mathbf{1}_{[0,i]}}(i) = \sup_{0 \leq f \leq 1} \overrightarrow{g_f}(i). \quad (25)$$

Moreover if $V_1 \geq 0$, then for all $i \in \mathbb{N}^*$

$$\partial g_{\mathbf{1}_i}(i) = \max_{0 \leq f \leq 1} |\partial g_f(i)|. \quad (26)$$

Proof. By replacing f by $1 - f$ if necessary

$$\sup_{0 \leq f \leq 1} |g_f(i)| = \sup_{0 \leq f \leq 1} g_f(i), \quad \sup_{0 \leq f \leq 1} |\partial g_f(i)| = \sup_{0 \leq f \leq 1} \partial g_f(i).$$

By Lemma 6.1,

$$g_f(i+1) = e_{i+1}^- \sum_{j=0}^i \pi(j) f(j) - e_i^+ \sum_{j=i+1}^{\infty} \pi(j) f(j) \leq e_{i+1}^- \sum_{j=0}^i f(j),$$

with equality for $f = \mathbf{1}_{[0,i]}$ which proves (25). On the other hand,

$$\partial g_j(i) = \pi_j ((e_{i-1}^+ - e_i^+) \mathbf{1}_{i \leq j-1} + (e_{i+1}^- + e_{i-1}^+) \mathbf{1}_{i=j} + (e_{i+1}^- - e_i^-) \mathbf{1}_{i \geq j+1}),$$

so by Lemma 6.2 the quantity $\partial g_j(i)$ is non-negative if and only if $i = j$. Hence, if f is a function on \mathbb{N} with values in $[0, 1]$,

$$\partial g_f(i) = \sum_{j=0}^{\infty} f(j) \partial g_j(i) \leq \partial g_i(i),$$

and there is equality if $f = \mathbf{1}_i$. This shows (26). \square

As a consequence, we have the following lemma of which Theorem 3.2 is a direct application.

Lemma 6.4 (Pointwise first Stein's factor for bounded functions). *If V_u is bounded from below by $\sigma(u)$ then for all $i \in \mathbb{N}$,*

$$\sup_{0 \leq f \leq 1} |g_f(i+1)| \leq \int_0^{\infty} e^{-\sigma(u)t} \mathbb{P}(X_{u,t}^i = i) dt.$$

Moreover if V_u is constant then the preceding inequality is in fact an equality.

Proof. By the equation (20), Theorem 2.1 and Lemma 6.3, and because $\partial_u \mathbf{1}_{[0,i]} = -\frac{1}{u(i)} \mathbf{1}_i$, we have for all function f such that $0 \leq f \leq 1$

$$|g_f(i+1)| \leq g_{\mathbf{1}_{[0,i]}}(i+1) = -u(i) \int_0^{\infty} P_{u,t}^{V_u}(\partial_u \mathbf{1}_{[0,i]}) = \int_0^{\infty} P_{u,t}^{V_u}(\mathbf{1}_i) \leq \int_0^{\infty} e^{-\sigma(u)t} \mathbb{P}(X_{u,t}^i = i) dt.$$

\square

We now state results for the second pointwise Stein factor.

Lemma 6.5 (Pointwise second Stein's factor for bounded functions).

- Under \mathbf{H}_1 , for all integer $i \in \mathbb{N}^*$, the quantity $\sup_{0 \leq f \leq 1} |\partial g_f(i)|$ is bounded by

$$\int_0^\infty e^{-\sigma(1, *u)t} \left(-\frac{u(i)}{u(i-1)} \mathbb{P}(X_{1, *u, t}^i = i-1) + 2\mathbb{P}(X_{1, *u, t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1, *u, t}^i = i+1) \right) dt.$$

- Under \mathbf{H}_2 , for all integer $i \in \mathbb{N}$, the quantity $\sup_{0 \leq f \leq 1} |\partial g_f(i+1)|$ is bounded by

$$\int_0^\infty e^{-\sigma(1, u)t} \left(-\frac{u(i)}{u(i-1)} \mathbb{P}(X_{1, u, t}^i = i-1) + 2\mathbb{P}(X_{1, u, t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1, u, t}^i = i+1) \right) dt.$$

Moreover, if the potential $V_{1, *u}$ (respectively $V_{1, u}$) is constant, then the first (respectively the second) upper bound is in fact an equality.

Proof. For every positive integer i , let $f_i = \mathbf{1}_i$. By the equation (21), Theorem 2.2 and Lemma 6.3, under \mathbf{H}_1 ,

$$\sup_{0 \leq f \leq 1} |\partial g_f(i)| = \partial g_{f_i}(i) = u(i) \partial_u g_{f_i}(i) \leq u(i) \int_0^\infty e^{-\sigma(1, *u)t} \mathbb{E} \left[\partial_u^* \partial f_i(\tilde{X}_t^i) \right] dt.$$

As $\partial_u^* \partial f_i = -\frac{1}{u(i-1)} \mathbf{1}_{i-1} + 2\frac{1}{u(i)} \mathbf{1}_i - \frac{1}{u(i+1)} \mathbf{1}_{i+1}$, we get the announced inequality.

Similarly the result under \mathbf{H}_2 derives from the equation (22), Theorem 2.5, Lemma 6.3 and the computation $-\partial_u \partial f_{i+1} = -\frac{1}{u(i-1)} \mathbf{1}_{i-1} + 2\frac{1}{u(i)} \mathbf{1}_i - \frac{1}{u(i+1)} \mathbf{1}_{i+1}$. \square

Theorem 3.3 and 3.4 are direct consequences of Lemma 6.5.

6.2 Approximation in Wasserstein distance.

In contrast with the first order in total variation distance, the bound of Theorem 3.6 does not require a preliminary bound on pointwise Stein's factor.

Proof of Theorem 3.6. By Theorem 2.1,

$$\begin{aligned} \left| \frac{1}{u} g_f(\cdot + 1) \right| &= |\partial_u h_f| = \left| \int_0^\infty \partial_u P_t f dt \right| \\ &= \left| \int_0^\infty \mathbb{E} \left[e^{-\int_0^t V_u(X_u, s) ds} \partial_u f(X_{u, t}) \right] dt \right| \leq \frac{1}{\sigma(u)} \|\partial_u f\|_\infty. \end{aligned}$$

Now to prove the sharpness if V_u is constant, it is enough to consider the map $f : x \mapsto -\sum_{k=1}^x u(x-1)$ for which the previous inequalities are in fact equalities. \square

Remark 6.6 (Variant of Theorem 3.6). *We can also derive an upper bound for*

$$\sup_{f \in \text{Lip}(d_u)} \|g_f/u\|_\infty,$$

*under the condition that V_{*u} is bounded by below, by using alternatively to equation (20) the equation*

$$g_f = -u \int_0^\infty \partial_u^* P_t f dt,$$

and Theorem 5.1 instead of Theorem 2.1.

For the second Stein factor, we begin by focusing on the pointwise quantity $\sup_{f \in \mathcal{F}} \partial_u g_f(i)$. For all $i \in \mathbb{N}$, let us introduce two functions ψ_i and Ψ_i defined for all $j \in \mathbb{N}$ as

$$\begin{aligned}\psi_i(j) &= \left(1 - \frac{u(j-1)}{u(j)}\right) \mathbf{1}_{j \leq i-1} + \left(1 + \frac{u(j-1)}{u(j)}\right) \mathbf{1}_{j=i} + \left(\frac{u(j-1)}{u(j)} - 1\right) \mathbf{1}_{j \geq i+1} \\ \Psi_i(j) &= \left(1 - \frac{u(j+1)}{u(j)}\right) \mathbf{1}_{j \leq i-1} + \left(1 + \frac{u(j+1)}{u(j)}\right) \mathbf{1}_{j=i} + \left(\frac{u(j+1)}{u(j)} - 1\right) \mathbf{1}_{j \geq i+1}.\end{aligned}$$

The following lemma allows to determine the functions that realize the supremum in the second pointwise Stein factor. This lemma is a generalization of a lemma of Barbour and Xia [2006], which addressed the case where $u = 1$ and $(\alpha(x), \beta(x))_{x \in \mathbb{N}} = (\lambda, x)_{x \in \mathbb{N}}$. Its proof depends on the already cited results of Brown and Xia [2001].

Lemma 6.7 (Argmax of the pointwise Stein's factor). *If $V_1 \geq 0$, then for all $i \in \mathbb{N}^*$*

$$\partial g_{\varphi_i} = \max_{f \in \text{Lip}(d_u)} |\partial g_f(i)|, \quad \varphi_i = -d_u(i, \cdot). \quad (27)$$

Proof. If f and \tilde{f} are two real-valued functions on \mathbb{N} , then $g_{f+\tilde{f}} = g_f + g_{\tilde{f}}$ and that if f is constant, then $g_f = 0$. As a consequence, by replacing f by $-f$ and $f - f(i)$ if necessary,

$$\sup_{f \in \text{Lip}(d_u)} |\partial g_f(i)| = \sup_{\substack{f \in \text{Lip}(d_u), \\ f(i)=0}} \partial g_f(i).$$

Recall that $g_j := g_{\mathbf{1}_j}$ for $j \in \mathbb{N}$. For all real-valued function f on \mathbb{N} ,

$$g_f = \sum_{j \in \mathbb{N}} f(j) g_j, \quad \partial g_f(i) = \sum_{j \in \mathbb{N}} f(j) \partial g_j(i), \quad i \in \mathbb{N}.$$

By Lemmas 6.1 and 6.2, if $f \in \text{Lip}(d_u)$ and $f(i) = 0$ then

$$\begin{aligned}\partial g_f(i) &= \sum_{j=0}^{\infty} f(j) \partial g_j(i) = (e_{i-1}^+ - e_i^+) \sum_{j \leq i-1} \pi_j f(j) + (e_{i+1}^- - e_i^-) \sum_{j \geq i+1} \pi_j f(j) \\ &\leq |\partial g_{\varphi_i}(i)| = (e_i^+ - e_{i-1}^+) \sum_{j \leq i-1} \pi_j d_u(i, j) + (e_i^- - e_{i+1}^-) \sum_{j \geq i+1} \pi_j d_u(i, j).\end{aligned}$$

□

Lemma 6.8 (Pointwise second Stein's factor for Lipschitz functions).

- Under \mathbf{H}_1 , for all integer $i \in \mathbb{N}^*$,

$$\sup_{f \in \text{Lip}(d_u)} |\partial_u g_f(i)| \leq \int_0^\infty e^{-\sigma(1, *u)t} \mathbb{E}[\psi_i(X_{1, *u, t}^i)] dt. \quad (28)$$

Moreover if $V_{1, *u}$ is constant then the preceding inequality is in fact an equality.

- Under \mathbf{H}_2 , for all integer $i \in \mathbb{N}$

$$\sup_{f \in \text{Lip}(d_u)} \left| \frac{1}{u(i)} \partial g_f(i+1) \right| \leq \int_0^\infty e^{-\sigma(1, u)t} \mathbb{E}[\Psi_i(X_{1, u, t}^i)] dt. \quad (29)$$

Moreover if $V_{1, u}$ is constant then the preceding inequality is in fact an equality.

Proof of Lemma 6.8. Let us assume that \mathbf{H}_1 holds true. By the equation (21), Theorem 2.2 and Lemma 6.7, for every positive integer i ,

$$\begin{aligned} \sup_{f \in \text{Lip}(d_u)} |\partial_u g_f(i)| &= \frac{1}{u(i)} \sup_{f \in \text{Lip}(d_u)} |\partial g_f(i)| = \frac{1}{u(i)} \partial g_{\varphi_i}(i) = \partial_u g_{\varphi_i}(i) = \int_0^\infty \partial_u^* \partial P_t \varphi_i dt \\ &= \int_0^\infty P_{1,*u,t}^{V_{1,*u}} (\partial_u^* \partial \varphi_i) dt \leq \int_0^\infty e^{-\sigma(1,*u)t} \mathbb{E} [\partial_u^* \partial \varphi_i(X_{1,*u,t}^i)] dt. \end{aligned}$$

It is easy to check that $\psi_i = \partial_u^* \partial \varphi_i$, which proves (28).

Now, if \mathbf{H}_2 holds true, by the equation (22), Theorem 2.5 and Lemma 6.7, for all integer i ,

$$\begin{aligned} \sup_{f \in \text{Lip}(d_u)} \left| \frac{1}{u(i)} \partial g_f(i+1) \right| &= \frac{1}{u(i)} \sup_{f \in \text{Lip}(d_u)} |\partial g_f(i+1)| = \frac{1}{u(i)} \partial g_{\varphi_{i+1}}(i+1) \\ &= - \int_0^\infty \partial_u \partial P_t \varphi_{i+1}(i) dt \\ &= - \int_0^\infty P_{1,u,t}^{V_{1,u}} \partial_u \partial \varphi_{i+1}(i) dt \end{aligned}$$

As $-\partial_u \partial \varphi_{i+1} = \Psi_i$, the equation (29) holds true. \square

We deduce from Lemma 6.8 both Theorem 3.7 and Theorem 3.8. We only give the proof of Theorem 3.7 because Theorem 3.8 is similar.

Proof of Theorem 3.7. First of all let us notice that for all function $f : \mathbb{N} \rightarrow \mathbb{R}$,

$$\|\partial_u f\|_\infty \leq \sup_{x \in \mathbb{N}} \left(1 + \frac{u(x+1)}{u(x)} \right) \|f/u\|_\infty, \quad \|\partial_u^* f\|_{\infty, \mathbb{N}^*} \leq \sup_{x \in \mathbb{N}^*} \left(1 + \frac{u(x-1)}{u(x)} \right) \|f/u\|_\infty. \quad (30)$$

Under \mathbf{H}_1 , as $\|\partial_u \varphi_i\|_\infty \leq 1$, it implies that

$$\sup_{x \in \mathbb{N}^*} |\partial_u \partial^* \varphi_i(x)| \leq \sup_{x \in \mathbb{N}^*} \left(1 + \frac{u(x-1)}{u(x)} \right).$$

Plugging this in the equation (28) yields the first upper bound of the theorem.

On the other hand, if $u(x) = q^x$ on \mathbb{N} with $q \geq 1$, then by using that $\mathbf{1}_{[0,i]} = 1 - \mathbf{1}_i - \mathbf{1}_{[i+1,\infty)}$, we write

$$\begin{aligned} \partial_u^* \partial \varphi_i(j) &= \left(1 - \frac{u(j-1)}{u(j)} \right) + 2 \frac{u(j-1)}{u(j)} \mathbf{1}_{j=i} + 2 \left(\frac{u(j-1)}{u(j)} - 1 \right) \mathbf{1}_{j \geq i+1} \\ &\leq 1 - \frac{1}{q} + 2 \frac{1}{q} \mathbf{1}_{j=i} \end{aligned}$$

which proves the second upper bound. \square

6.3 Approximation in Kolmogorov distance

Proof of Theorem 3.9. As one can see in Lemma 6.3, the function that realizes the maximum in the first Stein factor associated to bounded functions, $f = \mathbf{1}_{[0,i]}$, is also an element of the class of the half-line indicator functions. Hence without further analysis the analogous of Lemma 6.4 and Theorem 3.2 hold by replacing $\mathcal{F} = \{0 \leq f \leq 1\}$ by $\mathcal{F} = \{\mathbf{1}_{[0,m]}, m \in \mathbb{N}\}$. \square

For the second Stein factor, we begin by determining the argmax of the pointwise factor, as we did previously.

Lemma 6.9 (Argmax of the pointwise Stein factor). *For all $i \in \mathbb{N}$*

$$\max \left\{ -\partial g_{\mathbf{1}_{[0, i-1]}}(i), \partial g_{\mathbf{1}_{[0, i]}}(i) \right\} = \sup_{f=\mathbf{1}_{[0, m]}, m \in \mathbb{N}} |\partial g_f(i)|.$$

Proof. Let $f = \mathbf{1}_{[0, m]}$ for an integer m . By Lemma 6.1, if $m \leq i - 1$,

$$\partial g_g(i) = \sum_{j=0}^m \pi(j)(e_{i+1}^- - e_i^-).$$

Hence by Lemma 6.2,

$$|\partial g_f(i)| = -\partial g_g(i) = (e_i^- - e_{i+1}^-) \sum_{j=0}^m \pi(j),$$

so the maximum when m browses the interval $[0, i - 1]$ is attained in $m = i - 1$.

Now, if $m \geq i$, let us call $F = 1 - f = \mathbf{1}_{[m+1, \infty)}$. By the same lemmas,

$$|g_f(i)| = |g_F(i)| = |e_i^+ - e_{i-1}^+| \sum_{j=m+1}^{\infty} \pi(j) = (e_i^+ - e_{i-1}^+) \sum_{j=m+1}^{\infty} \pi(j) = g_f(i),$$

so the maximum when m browses the interval $[i, +\infty)$ is attained in $m = i$. \square

Lemma 6.10 (Second pointwise Stein's factor for indicator functions).

- Under \mathbf{H}_1 , for all integer $i \in \mathbb{N}^*$, the quantity $\sup_{f=\mathbf{1}_{[0, m]}, m \in \mathbb{N}} \partial g_f(i)$ is bounded by the maximum of

$$\int_0^\infty e^{-\sigma(1, *u)t} \left(\mathbb{P}(X_{1, *u, t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1, *u, t}^i = i-1) \right) dt$$

and

$$\int_0^\infty e^{-\sigma(1, *u)t} \left(\mathbb{P}(X_{1, *u, t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1, *u, t}^i = i+1) \right) dt.$$

- Under \mathbf{H}_2 , for all integer $i \in \mathbb{N}$, the quantity $\sup_{f=\mathbf{1}_{[0, m]}, m \in \mathbb{N}} \partial g_f(i+1)$ is bounded by the maximum of

$$\int_0^\infty e^{-\sigma(1, u)t} \left(\mathbb{P}(X_{1, u, t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1, u, t}^i = i-1) \right) dt$$

and

$$\int_0^\infty e^{-\sigma(1, u)t} \left(\mathbb{P}(X_{1, u, t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1, u, t}^i = i+1) \right) dt.$$

Moreover, if the potential $V_{1, *u}$ (respectively $V_{1, u}$) is constant, then the first (respectively the second) upper bound is in fact an equality.

Proof. If $f = \mathbf{1}_{[0, m]}$ then $\partial_u^* \partial f_m = \frac{1}{u(m)} \mathbf{1}_m - \frac{1}{u(m+1)} \mathbf{1}_{m+1}$. Under \mathbf{H}_1 , by equation (21) and Theorem 2.2,

$$\begin{aligned} -\partial g_{\mathbf{1}_{[0, i-1]}}(i) &= u(i) \int_0^\infty P_{1, *u, t}^{V_{1, *u}} \left(-\frac{1}{u(i-1)} \mathbf{1}_{i-1} + \frac{1}{u(i)} \mathbf{1}_i \right) dt \\ &\leq \int_0^\infty e^{-\sigma(1, *u)t} \left(\mathbb{P}(X_{1, *u, t}^i = i) - \frac{u(i)}{u(i-1)} \mathbb{P}(X_{1, *u, t}^i = i-1) \right) dt. \end{aligned}$$

Similarly,

$$\partial g_{1_{[0,i]}}(i) \leq \int_0^\infty e^{-\sigma(1,*u)t} \left(\mathbb{P}(X_{1,*u,t}^i = i) - \frac{u(i)}{u(i+1)} \mathbb{P}(X_{1,*u,t}^i = i+1) \right) dt.$$

We get the conclusion by Lemma 6.9. The proof is analogous under \mathbf{H}_2 , using this time equation (22) and Theorem 2.5. \square

Finally, Theorem 3.10 and 3.11 are simple consequences of the previous lemma.

7 Proof of Section 4

The second upper bound of Lemma 4.1 derives by classical arguments from Mehler's formula (15) and the following lemma.

Lemma 7.1 (Upper bound on differences of the pointwise probabilities of the Poisson distribution).

$$\sup_{x \in \mathbb{N}} |\mathcal{P}_\lambda(x) - \mathcal{P}_\lambda(x-1)| \leq 1 \wedge \frac{C}{\lambda}, \quad C := \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \leq 1. \quad (31)$$

Proof of Lemma 7.1. Set

$$q(\lambda, x) = \lambda |\mathcal{P}_\lambda(x) - \mathcal{P}_\lambda(x-1)|, \quad x \in \mathbb{N}, \quad \lambda > 0.$$

Let us show that

$$\sup_{x \in \mathbb{N}, \lambda > 0} q(\lambda, x) < +\infty.$$

Firstly,

$$q(\lambda, x) = \frac{\lambda^x e^{-\lambda}}{x!} |\lambda - x|, \quad x \in \mathbb{N}, \quad \lambda > 0.$$

We first deal with the case where $x \in \mathbb{N}^*$. By a formula of Robbins (Robbins [1955]), we know that for all $x \in \mathbb{N}^*$,

$$x! > \sqrt{2\pi x} x^x e^{-x + \frac{1}{12x}} \geq \sqrt{2\pi} e^{\frac{1}{2} \log x + x \log x - x}.$$

Hence, $q(\lambda, x) \leq \frac{1}{\sqrt{2\pi}} e^{f(\lambda, x)}$ with

$$\begin{aligned} f(\lambda, x) &= x - \lambda + \log |x - \lambda| - \frac{1}{2} \log x + x \log \frac{\lambda}{x}, \\ \partial_\lambda f(\lambda, x) &= -1 + \frac{x}{\lambda} + \frac{1}{\lambda - x}. \end{aligned}$$

In the sequel we derive upper bounds of $f(\lambda, x)$ on relevant subsets of $(0, \infty) \times [1, \infty)$. One finds that

$$\partial_\lambda f(\lambda, x) = 0 \Leftrightarrow (\lambda - x)(x - \lambda + \sqrt{\lambda})(x - \lambda - \sqrt{\lambda}) = 0.$$

Let us call $\lambda_1(x)$ the solution of the equation $x = \lambda + \sqrt{\lambda}$ and $\lambda_2(x)$ the solution of the equation $x = \lambda - \sqrt{\lambda}$. We have $0 < \lambda_1(x) < x < \lambda_2(x)$.

If $\lambda \leq x$, then at x fixed the function $f(\lambda, x)$ is increasing on $(0, \lambda_1(x)]$ and decreasing on $[\lambda_1(x), x]$. Hence,

$$\sup_{x \geq 1, 0 < \lambda \leq x} f(\lambda, x) = \sup_{x \geq 1} f(\lambda_1(x), x) = \sup_{\lambda > 0} f(\lambda, \lambda + \sqrt{\lambda}).$$

Moreover, using that $\forall u \geq 0, \log(1+u) \geq u - u^2/2$, we find that

$$\begin{aligned}
f(\lambda, \lambda + \sqrt{\lambda}) &= \sqrt{\lambda} + \frac{1}{2} \log \frac{\lambda}{\lambda + \sqrt{\lambda}} + (\lambda + \sqrt{\lambda}) \log \frac{\lambda}{\lambda + \sqrt{\lambda}} \\
&= \sqrt{\lambda} - (\lambda + \sqrt{\lambda} + \frac{1}{2}) \log \left(1 + \frac{1}{\sqrt{\lambda}} \right) \\
&\leq \sqrt{\lambda} - (\lambda + \sqrt{\lambda} + \frac{1}{2}) \left(\frac{1}{\sqrt{\lambda}} - \frac{1}{2\lambda} \right) \\
&= -\frac{1}{2} \left(1 - \frac{1}{2\lambda} \right).
\end{aligned} \tag{32}$$

Hence, if $\lambda \geq \frac{1}{2}$ then $f(\lambda, \lambda + \sqrt{\lambda}) \leq 0$. If $\lambda \leq \frac{1}{2}$, by going back up to the equation (32),

$$f(\lambda, \lambda + \sqrt{\lambda}) \leq \sqrt{\lambda} \leq \frac{1}{\sqrt{2}}.$$

At the end,

$$\sup_{x \in \mathbb{N}^*, 0 < \lambda \leq x} f(\lambda, x) \leq \frac{1}{\sqrt{2}}.$$

Let us call $C_1 := \frac{1}{\sqrt{2\pi}} e^{\frac{1}{\sqrt{2}}} \sim 0,8$.

Now let us deal with the case where $\lambda \geq x$. We apply a different strategy for small integers x as for large integers x . First of all, for all $x \in \mathbb{N}^*$ and for all $\lambda \geq x$,

$$q(\lambda, x) = \frac{1}{x!} e^{-\lambda} \lambda^x (\lambda - x) \leq \frac{1}{x!} e^{-\lambda} \lambda^{x+1}$$

and it is easy to see that at x fixed the maximum of the right-hand expression is attained at $\lambda = x + 1$. Hence

$$q(\lambda, x) \leq \frac{1}{x!} e^{-(x+1)} (x+1)^{x+1}.$$

Hence

$$\sup_{x \in \{1,2,3\}, \lambda \geq x} q(\lambda, x) \leq C_2 := \max_{x \in \{1,2,3\}} \frac{1}{x!} e^{-(x+1)} (x+1)^{x+1} \sim 0.7.$$

On the other hand, by the same reasoning as below, we find that

$$\sup_{x \in [4, +\infty), \lambda \geq x} f(\lambda, x) = \sup_{x \in [4, +\infty)} f(\lambda_2(x), x) = \sup_{\lambda \geq 4} f(\lambda, \lambda - \sqrt{\lambda}).$$

Now, for all $\lambda > 1$,

$$f(\lambda, \lambda - \sqrt{\lambda}) = -\sqrt{\lambda} - \left(\frac{1}{2} + \lambda - \sqrt{\lambda} \right) \log \left(1 - \frac{1}{\sqrt{\lambda}} \right).$$

We use that $\forall u \in [0, \frac{1}{2}], -\log(1-u) \leq u + u^2$. As $\lambda \geq 4$ implies $\frac{1}{\sqrt{\lambda}} \leq \frac{1}{2}$,

$$\begin{aligned}
f(\lambda, \lambda - \sqrt{\lambda}) &\leq -\sqrt{\lambda} + \left(\frac{1}{2} + \lambda - \sqrt{\lambda} \right) \left(\frac{1}{\sqrt{\lambda}} + \frac{1}{\lambda} \right) \\
&= -\frac{1}{2\sqrt{\lambda}} \left(1 - \frac{1}{\sqrt{\lambda}} \right) \\
&\leq 0.
\end{aligned}$$

At the end,

$$\sup_{\lambda \geq x \geq 4} f(\lambda, x) \leq C_3 := \frac{1}{\sqrt{2\pi}} \sim 0.4.$$

It remains the case where $x = 0$, for which it is trivial to see that

$$q(\lambda, 0) = \lambda e^{-\lambda} \leq C_4 = e^{-1}, \quad \lambda > 0.$$

The final result follows with $C = \max\{C_1, C_2, C_3, C_4\} = C_1$. □

References

- Joseph Abate, Masaaki Kijima, and Ward Whitt. Decompositions of the $M/M/1$ transition function. *Queueing Systems Theory Appl.*, 9(3):323–336, 1991.
- François Baccelli and William A. Massey. A sample path analysis of the $M/M/1$ queue. *J. Appl. Probab.*, 26(2):418–422, 1989.
- D. Bakry and Michel Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.
- A. D. Barbour and T. C. Brown. Stein’s method and point process approximation. *Stochastic Process. Appl.*, 43(1):9–31, 1992.
- A. D. Barbour and Aihua Xia. On Stein’s factors for Poisson approximation in Wasserstein distance. *Bernoulli*, 12(6):943–954, 2006.
- A. D. Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1992. Oxford Science Publications.
- A. D. Barbour, H. L. Gan, and A. Xia. Stein factors for negative binomial approximation in Wasserstein distance. *Bernoulli*, 21(2):1002–1013, 2015.
- Michel Bonnefont and Aldéric Joulin. Intertwining relations for one-dimensional diffusions and application to functional inequalities. *Potential Anal.*, 41(4):1005–1031, 2014.
- Timothy C. Brown and M. J. Phillips. Negative binomial approximation with Stein’s method. *Methodol. Comput. Appl. Probab.*, 1(4):407–421, 1999.
- Timothy C. Brown and Aihua Xia. Stein’s method and birth-death processes. *Ann. Probab.*, 29(3):1373–1403, 2001.
- Djalil Chafaï and Aldéric Joulin. Intertwining and commutation relations for birth-death processes. *Bernoulli*, 19(5A):1855–1879, 2013.
- Louis H. Y. Chen. Poisson approximation for dependent trials. *Ann. Probability*, 3(3):534–545, 1975.
- Mu-Fa Chen. Estimation of spectral gap for Markov chains. *Acta Math. Sinica (N.S.)*, 12(4):337–360, 1996.
- Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.
- Peter Eichelsbacher and Gesine Reinert. Stein’s method for discrete Gibbs measures. *Ann. Appl. Probab.*, 18(4):1588–1618, 2008.

- David G. Kendall. On some modes of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, 35:6–15, 1948.
- Torgny Lindvall. *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.
- S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.
- Erol A. Peköz. Stein's method for geometric approximation. *J. Appl. Probab.*, 33(3):707–713, 1996.
- Erol A. Peköz, Adrian Röllin, and Nathan Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632, 2013.
- M. J. Phillips. *Stochastic process approximation and network applications*. PhD thesis, Univ. Melbourne, 1996.
- Emmanuel Rio. Distances minimales et distances idéales. *C. R. Acad. Sci. Paris Sér. I Math.*, 326(9):1127–1130, 1998.
- Herbert Robbins. A remark on Stirling's formula. *Amer. Math. Monthly*, 62:26–29, 1955.
- Charles Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602. Univ. California Press, Berkeley, Calif., 1972.
- A. Szulga. On minimal metrics in the space of random variables. *Teor. Veroyatnost. i Primenen.*, 27(2):401–405, 1982.
- V. M. Zolotarev. Metric distances in spaces of random variables and their distributions. *Mat. Sb. (N.S.)*, 101(143)(3):416–454, 456, 1976.